ARD
CORPORATION

FINAL REPORT

SIGNAL FEATURE ANALYSIS USING NEURAL NETWORKS & PSYCHOACOUSTICS

CONTRACT NO. N00014-90-C-0228

SUBMITTED TO:

OFFICE OF NAVAL RESEARCH

800 North Quincy Street
Arlington, Virginia 22217-5000

MAY, 1993

9151 Rumsey Road • Columbia, Maryland 21045 • (410) 997-5600 • Telefax: (410) 997-3902

EXECUTIVE SUMMARY

GOALS

The primary goal of this research was to compare the processing strategies used by human subjects and neural networks in classifying acoustic signals. Secondary goals were to compare subjects with and without sonar training and to investigate the effects on the neural networks of adding noise to the acoustic signals.

SIGNALS

The initial signal set was designed to provide a challenging classification task. The set was created by placing hollow metal acoustic targets on a sandy bottom in a large tank of water, insonifying them with a sonar signal, and collecting the reflected energy. The bottom environment was selected to provide reverberation to obscure the return from the target, making the classification task more difficult. For reference, signals were also collected from the targets suspended in the water column. The signal sets incorporated parameters by which the resulting signal classes differed: Material (Brass or Steel), Thickness (5% or 10% of outside diameter), and Angle (90°, 45°, or 0° to the insonifying beam). Subjects and networks were asked to classify the signals by these parameters.

Pilot experiments indicated that the classification of the underwater signals was very difficult, so a third signal set was created. The original targets were physically struck and the resulting vibrations were recorded. This signal set, denoted as "Air" signals, lacked the parameter of angle but added the parameter of striker (metal, plastic, and wood).

CLASSIFICATION EXPERIMENTS

After considerable signal processing to make the underwater signals audible, human subjects classified signals from each set in a series of experiments. Subjects were asked to identify each parameter of the signals separately. Over several sessions subjects received feedback in which the correct class of the current signal was revealed, then took a final session without feedback. Two groups of subjects were tested. The primary group, upon whose results signal processing

strategies were derived, was made up of Navy sonar personnel. The other group consisted of college students.

EXPERIMENT RESULTS

Experiments with the Bottom and Free-field signals revealed that these classification tasks were very difficult. In both cases only Angle was classified at levels above chance. Results from the Bottom classification experiment, however, indicated that the Navy subjects classified Angle correctly at a level significantly higher than that of the students.

The Air signal set was less difficult to classify. Both Navy and student subjects performed at levels higher than chance on all parameters of the Air signal set. Striker was the most difficult parameter to classify. Faced with a classification task of reasonable difficulty, the Navy subjects performed significantly higher than the student subjects by several measures.

MULTIDIMENSIONAL SCALING

During the classification experiments both the correct and incorrect responses of the subjects were recorded. These became raw data for confusion matrices which described how often a subject confused the class of a signal presented in the experiment with every other signal class. Multidimensional scaling was used to create a geometrical model of this data, in which the distance between signal classes is related to the degree of confusion between the classes. Only the best Navy subjects were modeled in this manner. The scaling solutions provided the dimensions which were taken to reflect subject strategies.

NEURAL NETWORKS

Backpropagation networks were trained to classify the preprocessed signals using signal transforms in both the time and frequency domains. Integrator gateway networks were also trained, using frequency information taken from a sliding window over the duration of the signals.

For each signal set, backpropagation networks were developed using a training set which consisted of half of the available signals, and a validation set made up of the other half of the available signals. The networks did not see the validation set while learning was enabled. As training

progressed the validation set was periodically presented with learning disabled, and the network weights that produced the highest performance on the validation set were recorded. Performance results are based on testing these weights with the validation set. Networks were trained using several different numbers of hidden nodes to evaluate the best architecture. Performance results are summarized in Table ES-1.

		Number of Hidden Nodes			
		0	2	4	6
Frequency	Free Field	Excellent	Difficulty on Angle	Excellent	Excellent
	Bottom	Excellent	Difficulty on Material	Excellent	Excellent
	Air	Excellent	Difficulty on Striker	Excellent	Excellent
Time	Free Field	Excellent	Difficulty on Material, Thickness	Excellent	Excellent
	Bottom	Excellent	Difficulty on Material, Thickness	Excellent	Excellent
	Air	Difficulty on Striker	Difficulty on Material, Striker	Difficulty on Striker	Difficulty on Striker

Table ES-1 Classification performance of backpropagation neural networks. "Excellent" indicates performance from 95 to 100%. "Difficulty" indicates performance from 60 to 90% on the indicated parameter, excellent on other parameters.

In the frequency domain, all networks performed very well except those with two hidden nodes. Within each signal set, networks with two hidden nodes had difficulty with one parameter, while performing well on the other two parameters. This is attributed to the relative lack of free parameters (weights) in comparison to networks with 0, 4, or 6 hidden nodes. While excellent performance without a hidden layer indicates that the problem may be linear, there were a large number of parameters available to these networks since all inputs were connected to all outputs.

The relatively poor performance of networks with two hidden nodes persisted in the time domain. The parameters that were troublesome changed for Free-field and Bottom networks, giving some indication as to which transforms of the signals carry the most information about which parameters. The Air networks did not perform as well on the Striker parameter when using time domain input. Human subjects also had the most difficulty with the Striker parameter.

Neural networks performed well on the classification task when properly configured and trained. They achieved high performance using signal data in either time or frequency domain. Air networks showed a preference for data in frequency domain based on relative performances. Four hidden nodes was generally the best architecture to balance high performance and a reasonably small number of free parameters in the network.

EFFECTS OF ARTIFICIAL NOISE

These networks were tested with signals to which increasing levels of random noise were added. As the signal-to-noise ratio (SNR) decreased so did the classification performance, although the networks were somewhat robust to reasonable noise levels. Performance fell off gradually. When comparable networks were trained using signals to which noise was added, the resulting networks were almost always more robust to noise than networks trained without noise added to the inputs.

INTEGRATOR GATEWAY NETWORKS

Integrator Gateway Networks (IGN) were also successful at the classification task. These networks take input in the form of frequency information from a series of windows over the duration of the signal. Each window is applied to the network until the entire signal has been applied. IGNs use a complex architecture to record and process this data. These networks were trained with Bottom and Air signals.

IGNs trained with Air signals performed perfectly on Material and Thickness, and well on Striker. Bottom IGNs performed just above chance on Material and Thickness, and rather well on Angle. In both cases the networks' relative performances are the same as those of most subjects. When the confusion data from a Bottom IGN was scaled, the resulting dimensions matched those of the human subjects.

TOOLS FOR MODELING DIMENSIONS

Several measures of the signals were computed in order to model the human dimensions created by scaling. In the frequency domain the spectrum can be viewed as a probability density function. From that premise measures such as mean frequency and standard deviation of the frequency distribution were computed. Two measures in the time domain were computed by fitting an exponential to the envelope of the signals. Finally, each Air signal was fit with a series of decaying sine waves, which were characterized by several parameters each.

A number of these signal measures were highly correlated with human scaling dimensions. These correlations were assumed to indicate that the signal measure is a reasonable model of the signal processing on that dimension, lacking any means of directly measuring the processing of the subjects. In addition to the signal measures, every human dimension was also correlated to two or more neural network hidden nodes. That is, the activations generated at the hidden node for each signal class closely resembled the placement of the signal classes on a scaling dimension. The processing strategies of correlated hidden nodes was explored. Certain dimensions are also correlated between scaling solutions, and for this reason dimensions are often analyzed in pairs. The results of these analyses are summarized in Figures ES-1 and ES-2.

NETWORK HIDDEN NODES AND DIMENSIONS

Neural network hidden nodes often applied the same strategies as the subjects on particular dimensions. An example is the set of relationships among the first scaling dimension of the top three subjects ("Best") on Bottom signals, the first dimension of the single best subject ("N6"), and two correlated time domain hidden nodes. The subjects differentiated 90° signals from other signals on this dimension using the large transient characteristic of 90° signals. The hidden nodes applied the same strategy. Correlated nodes trained with frequency domain data applied a strategy which took advantage of a signal feature closely related to the transient.

A second example of subject and network parallel strategies is found on the first dimension of the Best scaling solution for Air signals and the first dimension of the N4 solution. These subjects were sensitive to differences in the rates of decay of the ringing portions of the signals, and to the highly related frequency domain feature of standard deviation. Two hidden nodes in the time

Scaling Dimension	Role of Dimension	Human Processing	Network Hidden Node	Role of Node	Node Processing
Bottom Best 1st and Bottom N4 1st	Discriminate 90° signals from others	High amplitude transient in 90° signals, absent in others	Time Hidden 3	Activate for 90° signals only	Sensitive to the high amplitude transient in 90° signals
			Time Noise Hidden 4	Activate for 90° signals and 5% 0° signals	Sensitive to high amplitude transient and later small transient characteristic of 5% 0° signals
			Frequency Hidden 1	Activate for 90° signals only	Detect 90° signals by their increased energy at the insonifying frequency, a feature related to the time domain transient
			Frequency Noise Hidden 1	Activate for 90° signals and 10% 45° signals	Sensitive to energy in a small low frequency band

Figure ES-1 Summary of Subject and Neural Network Hidden Node Processing for Bottom Signals.

Scaling Dimension	Role of Dimension	Human Processing	Network Hidden Node	Role of Node	Node Processing
Air Best 1st and Air N4 1st	Discriminate signals by Thickness	Rate of decay of signal in time domain, and associated frequency domain effects including standard deviation of frequency spectrum and frequency of ringing portion of signal.	Time Noise Hidden 2	Approximate sorting of signals by Thickness	Simple weights, high activation for fast decay
			Time Hidden 3	Approximate sorting of signals by Thickness; used at output layer to detect 5%, reject 10%	Complex weights, high activation for fast decay, 2 other strategies
			Frequency Hidden 2, Frequency Noise Hidden 2	Perfect sorting of signals by Thickness; used at output layer to detect 10%, reject 5%	Same weight values; high activation for signals with wider frequency distribution and higher frequency ringing portions
Air Best 2nd and Air N4 3rd	Separate signals by Material, with the Brass 5% group at one extreme and Steel Metal Striker at the other	Subjects arranged signals largely according to mean frequency content	Time Noise Hidden 3	Differentiate the Brass 5% signals from all others	High activation only for very long, high amplitude signals (the Brass 5% group); simple weights
			Time Hidden 2		High activation for all signals except very long, high amplitude (Brass 5%); complex weights
			Frequency Noise Hidden 3	Differentiate Steel Metal and Steel Plasitc	Activates for specific low frequencies suppresses for specific high frequencies
N4 2nd	Differentiate the Metal Striker signals from others	Initial amplitude of the most persistent sine wave component of signal	Frequency Hidden 4, Frequency Noise Hidden 4	Detect Metal Striker signals	Same weight values; low activation for high energy at low frequency (characteristic of Plastic and Wood Striker signals); high activation otherwise; special case for one Metal Striker signal
Best 3rd	Sort signal by Material (more accurate than Best 2nd)	Frequency domain; lack of low frequencies, frequency of ringing portion	Frequency Hidden 1, Frequency Noise Hidden 1	Detect Steel signals	Same weight values; sensitve to high frequency peaks in Steel, low frequency peaks in Brass 5%, and various peaks in Brass 10%

Figure ES-2 Summary of Subject and Neural Network Hidden Node Processing for Air Signals.

domain applied a processing strategy which measured the rate of decay. In addition, two hidden nodes in the frequency domain were sensitive to differences in standard deviations.

In the time domain, the nodes with the highest level of similarity to the dimension model had been trained using noisy inputs. These nodes employed virtually the same strategies are their human counterparts. When a correlated node had been trained without noisy inputs, it employed a more complex but clearly related strategy. The first dimensions of the Air scaling solutions provide an example. Nodes trained with frequency domain data usually showed no difference in strategies between those nodes trained with and without noise. The strategies, however, bore close resemblance to those of the correlated dimensions.

Some dimensions appeared to reflect subject strategies exclusive to a single domain. Network nodes from the other domain were nevertheless highly correlated. This can be seen in the two time domain hidden nodes which are correlated with the first dimensions of the Air scaling solution. Such a capability might be suggestive of strategies that the subjects could employ, particularly subjects who have not learned to extract all possible information from a signal.

SUMMARY

The primary goal of the project was achieved by comparing the acoustic processing strategies of subjects and networks. Networks usually developed essentially the same strategies as subjects when given signals in the proper domain. When the signals used to train a network were in the opposite domain of the strategy used by subjects, the network usually developed a related strategy. A secondary goal was to compare the classification performances of subjects who were and were not trained in sonar. Subjects trained in sonar were better classifiers in tasks of moderate difficulty. Another goal was to evaluate the effects of low SNR signals on the networks. Networks were made more robust to noise by training with corrupted signals.

EXTENSIONS

Within the current signal set, several logical extensions of the research may make sense. One might be interested in the weight structure of a network trained to produce the same output as that of a subject attempting to classify the signals. Differences between high and low performers could be investigated in this manner, as well as differences between various signal input transforms.

Explanations of the dimensions analyzed in this effort might also be forthcoming from the weight structures of networks trained to replicate the dimensions. Given their capability to learn signal features networks might be also be explored as intelligent automated assistants to sonar operators, scanning large amounts of data for certain features.

The human data has also not been fully tapped. Dimensions were derived only from top Navy performers. Differences in processing strategies between high and low performers, and Navy and student subjects, may be of interest. Finally, the techniques of the research should be applied to data more in keeping with the Navy subjects' typical acoustic processing tasks.

1.0 MOTIVATION

Both people and neural networks are often very good classifiers of acoustic signals into their classes of origin. Networks, in fact, often outperform people on signals of moderate to high complexity. People are assumed to apply certain signal processing techniques, in the context of the brain, to achieve a high level of performance on such tasks. Networks learn these classifications through application of examples and modification of the network's weight structure. The completed weight structure embodies the techniques by which the network accomplishes the classification task.

Neither human nor network processing is necessarily easy to describe when a task of sufficient complexity is performed. Since the network encodes its processing strategy on weights which are accessible, we are interested in means of analyzing those weights to derive the underlying processing strategies. Unfortunately, we cannot perform the same analysis of human processing strategies by looking at the analogous, physiological processes. Human processing must be inferred through analysis of data derived during the classification process.

The intent of the research described here is to derive the strategies of subjects asked to perform a set of classification tasks, and to compare those strategies to the strategies of neural networks performing the same tasks. Strategies of the human subjects were derived using multidimensional scaling techniques which convert data concerning the confusions subjects' experience during the classification task into a form which describes the relationships among the signals the subjects were attempting to classify.

Networks are often performing too well to provide such data, but their weight structures are immediately accessible. They are analyzed by locating those elements of a network which most closely recreate the relationships among the signals found by the multidimensional scaling process, observing the local weights and their relationships to other parts of the network, and applying signals from various classes and observing the local reaction of the network.

Several other objectives emerge from this main objective. The selection of signal sets is vital to the ensuing classification tasks, and three different sets are employed here which provide tasks of varying complexity. Human subjects are taken from two groups, in order to compare the performances of subjects with and without sonar training and to derive strategies from the highest performing trained subjects. The effects of obscuring the signals presented to networks with

artificial noise is of interest to judge the effect on performance, and more importantly, on the strategies developed by the networks. Thus this research focused on networks and humans classifying acoustic signals, and the analysis of their performance and strategies.

2.0 METHODOLOGY OVERVIEW

The goals of the project were to examine and identify the strategies used by human listeners and by neural networks to classify a challenging signal set, and to compare those strategies. The methodology applied to reach these goals is described below.

2.1 SIGNALS

The acoustic signal set was the basis for all classification tasks. Its design was a collaborative effort between ARD, Dr. Douglas Todoroff, and Dr. James Howard. A degree of difficulty was sought to provide a reasonable challenge to both subjects and networks. The strategies employed to accomplish a challenging task were expected to be of greater interest than those which would result from an easier task. A source of reverberation was sought to complicate the classification task. To this end the acoustic targets were placed on a sandy bottom. The bottom provided a reflection of the insonifying pulse, and also presumably altered the echo from the target from its "free-field" condition (suspended in the water column).

Signals were collected in a Navy laboratory under the supervision of Dr. Todoroff. The targets and collection scenarios were varied to produce three parameters by which the resulting signals varied: material of the target, thickness of the target, and angle between the axis of insonification and the axis of the target. Free-field signals were collected in addition to Bottom signals to provide a reference standard. As detailed in later sections, the underwater signals proved more difficult to classify than was ideal for the purpose of deriving strategies, so a third signal set was collected. This set consisted of acoustic signals generated by striking the targets manually with various materials. This set was referred to as the "air" set since it was not collected underwater. The resulting acoustic events proved appropriately difficult for human subjects to classify, and subsequent analyses were conducted on both Bottom and Air signals and the corresponding test results.

2.2 CLASSIFICATION EXPERIMENT

Data on human classification strategies were derived from experiments in which the subjects classified the signals from the three signal sets. After listening to a signal, the subject was asked to select the Material, Thickness, and Angle (or, in the case of the Air signals, Striker) of the target

from which that signal was created. During some sessions of the experiment feedback was provided so that the subjects could learn the correct classifications. For each signal presented, both the correct and actual responses were recorded. In addition to the separate parameters, the performance on all parameters simultaneously was of interest and was derived from the stored data.

In collaboration with Dr. David Kobus, a set of Navy sonar personnel was used as the primary subject group. For comparison a set of college students was also tested. Their performances are compared in Section 6. Although the tasks in these experiments did not necessarily resemble the sonar tasks that the navy personnel are trained for, using these subjects allowed us to compare their performances to those of subjects without a particular professional background in acoustic tasks. Although all hearing people have experience in processing acoustic information and making classifications based on acoustic data, the navy subjects may be better prepared to perform specific tasks based on this data by virtue of professional training and experience.

2.3 SCALING AND DIMENSIONS

The data generated during the classification experiments consisted of the subjects' judgments of the material, thickness, and angle/striker parameters for each signal presented. When such data is compared to the actual values of those parameters for the given signal, a confusion matrix results. The confusion matrix quantifies the degree to which any pair of signals is confused in the classification task. It is assumed that a pair of signals frequently confused by the subject sounds similar to the subject, and that the confusion data measure the degree of similarity.

With similarity data available, multidimensional scaling became an attractive means of modeling the subjects' responses. By this technique the similarity data were used to place the signals in a three-dimensional space in such a manner that the distances between signal pairs corresponded to the similarity judgments of the subjects for the pairs. The scaling technique also provided the individual dimensions on which the signals were placed. These dimensions are assumed to correspond to processing methods or strategies used by the subjects in performing the signal classifications.

2.4 ACOUSTIC MEASURES OF THE SIGNALS

To model these dimensions, and presumably the underlying strategies, several techniques were used to characterize the signals. These techniques, which ranged from finding the mean frequency content of the signals to fitting exponentially decaying curves to them, generated scalar measures of the signals using acoustic information. In many instances these measures were highly correlated with the values of the signals on the scaling dimensions, suggesting that the given measure was related to the signal processing strategy employed by the subject and represented on the correlated dimension.

2.5 NEURAL NETWORKS

In addition to the physical measures employed to model the dimensions, neural networks were employed to classify the signals. The networks' classification performances were compared to those of the subjects to reveal certain similarities and differences. Some networks were trained with time domain data, some with frequency domain data, and some with a combination of time and frequency data. The networks were trained with and without the addition of random noise to their inputs, which resulted in remarkable differences in network performance and in the structure of the resulting network weights (and thereby the strategies used by the networks to perform the classifications).

The network weights provided the means by which the networks' strategies were compared to the subjects' strategies. A subset of the trained network nodes gave output activations which were highly correlated with the human scaling dimensions. These network nodes were reaching the same 'conclusions' about the signals as did the subjects, at least as indicated by the scaling dimensions. It was therefore of considerable interest how the correlated nodes went about assigning activations to the various signals. These issues are explored in Section 10 by observation of the weights, by application of the signals to the nodes, and by comparison of the intermediate results of the nodes for various signals.

The methods used to accomplish the tasks and analyses set forth in this section, and the results of those analyses, are described in detail in the remainder of the report.

3.0 SIGNALS

The signal set was an extension of a signal set used in previous acoustic research. The design of this set was produced in consultation with Dr. Doug Todoroff of the Naval Coastal Systems Center (NCSC) in Panama City, Florida. The major departure from the earlier research was to place the targets on a sandy bottom to introduce a reverberation component to the signal set. These "Bottom" signals became the early centerpiece of the study. Signals were also collected from the same targets as they hung from monofilament in the water column of the same collection tank. These "Free-field" signals did not suffer the complexities of the bottom reflection or the effect of the bottom on the return from the target. These signals were meant to be the control set with which to judge the effects of the bottom reverberation on the target returns. As detailed in subsequent sections, poor initial subject performance on the Bottom set led to the collection of a third signal set. This set consisted of sounds produced when the targets were struck as the targets hung from monofilament, in air, by strikers made of various materials.

For the signal sets four acoustic targets were designed and constructed. Three separate sets of acoustic signals were generated from these targets. Two signal sets, consisting of underwater Free-field and Bottom reflection returns, were collected in laboratory facilities at NCSC. The third set, containing sounds from targets manually struck using various materials, was collected in a sound-attenuated laboratory at the Catholic University of America. For each of the sets the parameters of Material and Thickness were varied. The third parameter varied was either the angle of insonification, in the case of the Free-field and Bottom sets, or the Striker, in the case of the "Air" signals. The Material parameter had been identified in conversations with Dr. Todoroff and Dr. Howard as an extension of the complexity found in the signals used in our previous acoustic classification work. The Thickness parameter is a standard in mine classification work and was also used in our previous project.

3.1 FREE-FIELD AND BOTTOM SIGNALS

All of the signal sets were generated using four targets which were cylindrical, enclosed, hollow, and metal. They were constructed as steel and brass cylinders with rounded end caps, and measured four inches in length by 3/4 inches in diameter. As well as having different materials, the targets had two shell thicknesses which were measured as a percentage of outside diameter. For each Material two targets were made, one at 10% (called "Thick") and the other at 5% ("Thin")

of the shell diameter. For the Free-field and Bottom signals, the targets' angles relative to the transducer were also varied. The angles used were 90°, 45°, and 0°, where 90° was the broadside orientation, 0° was an end-on perspective, and 45° was in between the two. The combination of the parameters of two Materials, two Thicknesses, and three Angles produced 12 signal classes.

The Free-field and Bottom signals were active sonar returns generated by insonifying targets in a 10'x10'x7' tank filled with water. The tank and target setups are illustrated in Figures 3.1-1 and 3.1-2. The Free-field signals were so named because the targets were suspended in the tank by monofilament and were of sufficient distance from the walls and floor to avoid interfering reverberations. The Bottom reflection signals were a product of targets laid on a smooth sand surface so that the target energy collected was embedded in reverberation from the sand. The Air signals were created by hanging each target by monofilament from horizontal crosspieces on a vertical metal stand.

For the Free-field and Bottom cases the insonifying signals were generated at 200, 400, 600 and 800 kHz. A set of sinusoids of varying numbers of cycles were produced for each of the frequencies. As the targets were insonified under the various conditions their reflections were collected along the axis of insonification by a receiver. For the Free-field conditions the transducer/receiver pair and the target hung on a line parallel to the floor of the tank. The Bottom condition required that the transducer/receiver pair be angled toward the target on the sand, and a grazing angle of 45° from the floor of the tank was used. Once a signal's return energy passed through the receiver it was fed through a preamplifier and filter, and captured by a digital oscilloscope onto a personal computer. The hardware specifications for both conditions are detailed in Figure 3.1-3.

During the collection process settings on several of the hardware components were adjusted to maximize the quality of the signal being captured. For each group of signals from the same condition, the oscilloscope cursor, which controlled the points that were digitized, was adjusted to include all of the energy from the signals in the 2048 point window. A filter with choices for high-pass and low-pass settings was adjusted each time the insonifying frequency changed. The high-pass filter was always set at 100 kHz, but the low-pass filter was set according to the frequency of the insonifying pulse. For instance, it was set to 400 kHz for a 200 kHz pulse and to its highest option of 1 MHz for the 600 kHz and 800 kHz sinusoids. The combination of a separate preamplifier and the voltage scale on the oscilloscope controlled the relative amplitudes of

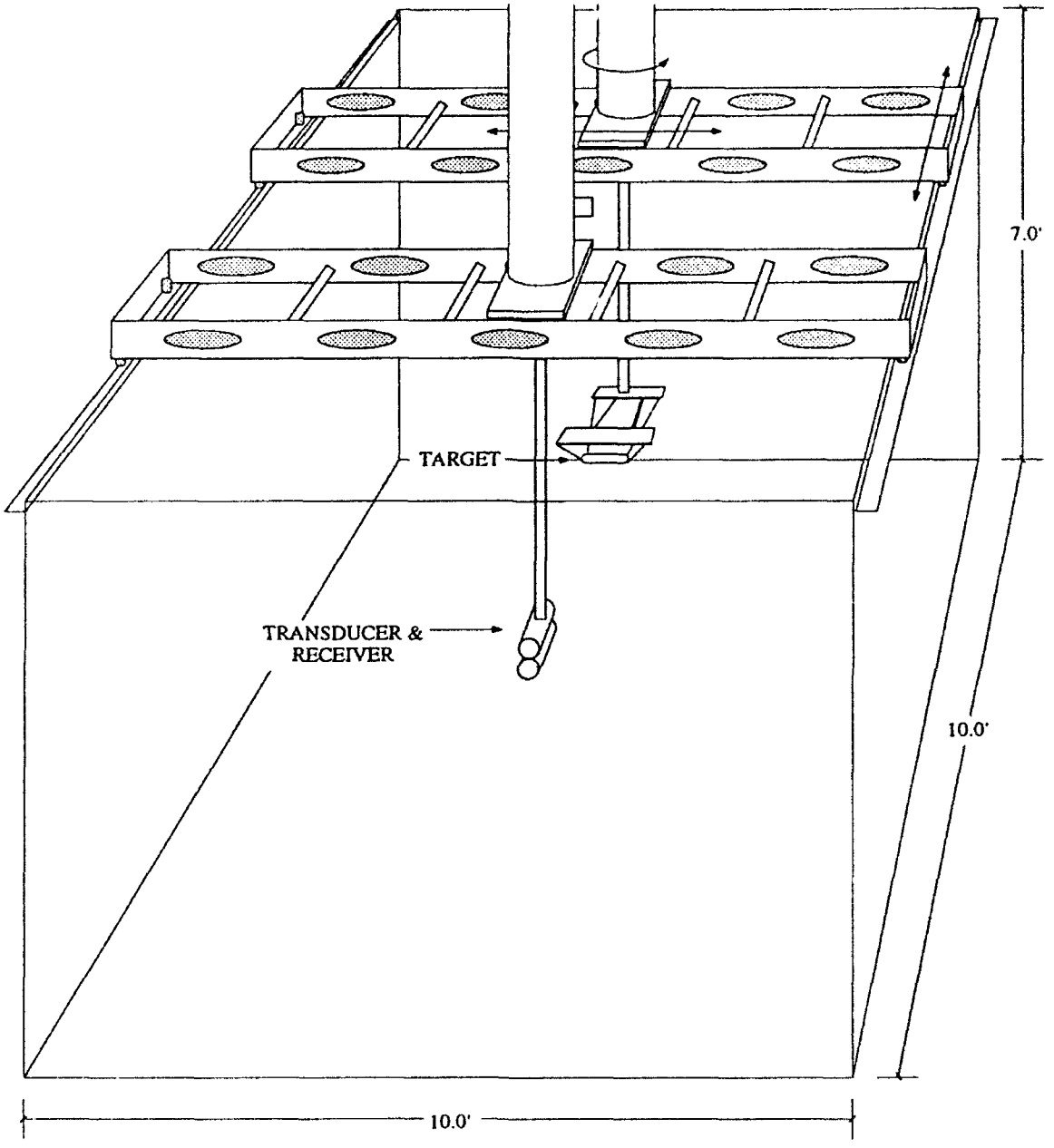


Figure 3.1-1 Tank Set-up for Free-Field Signal Collection

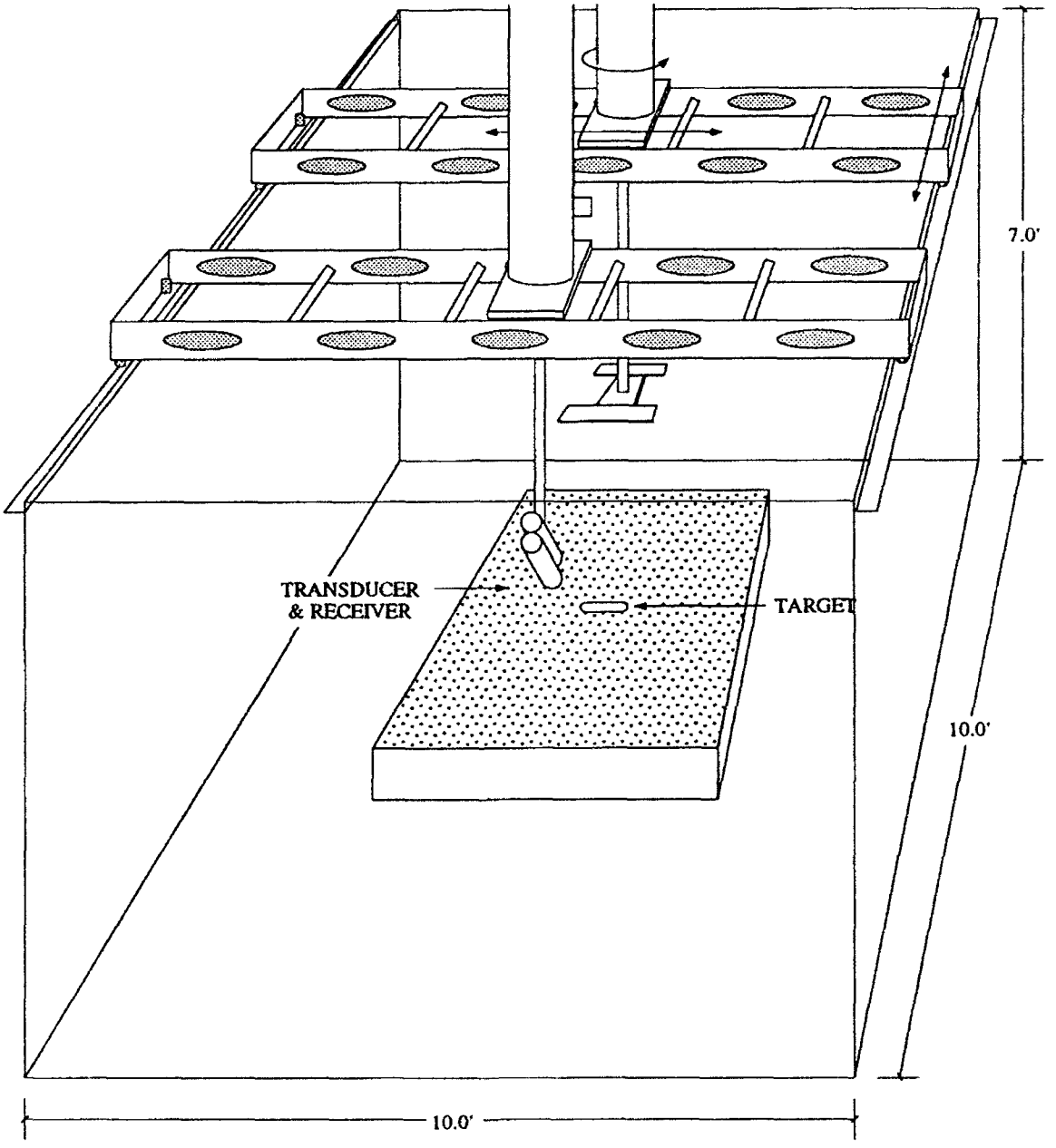
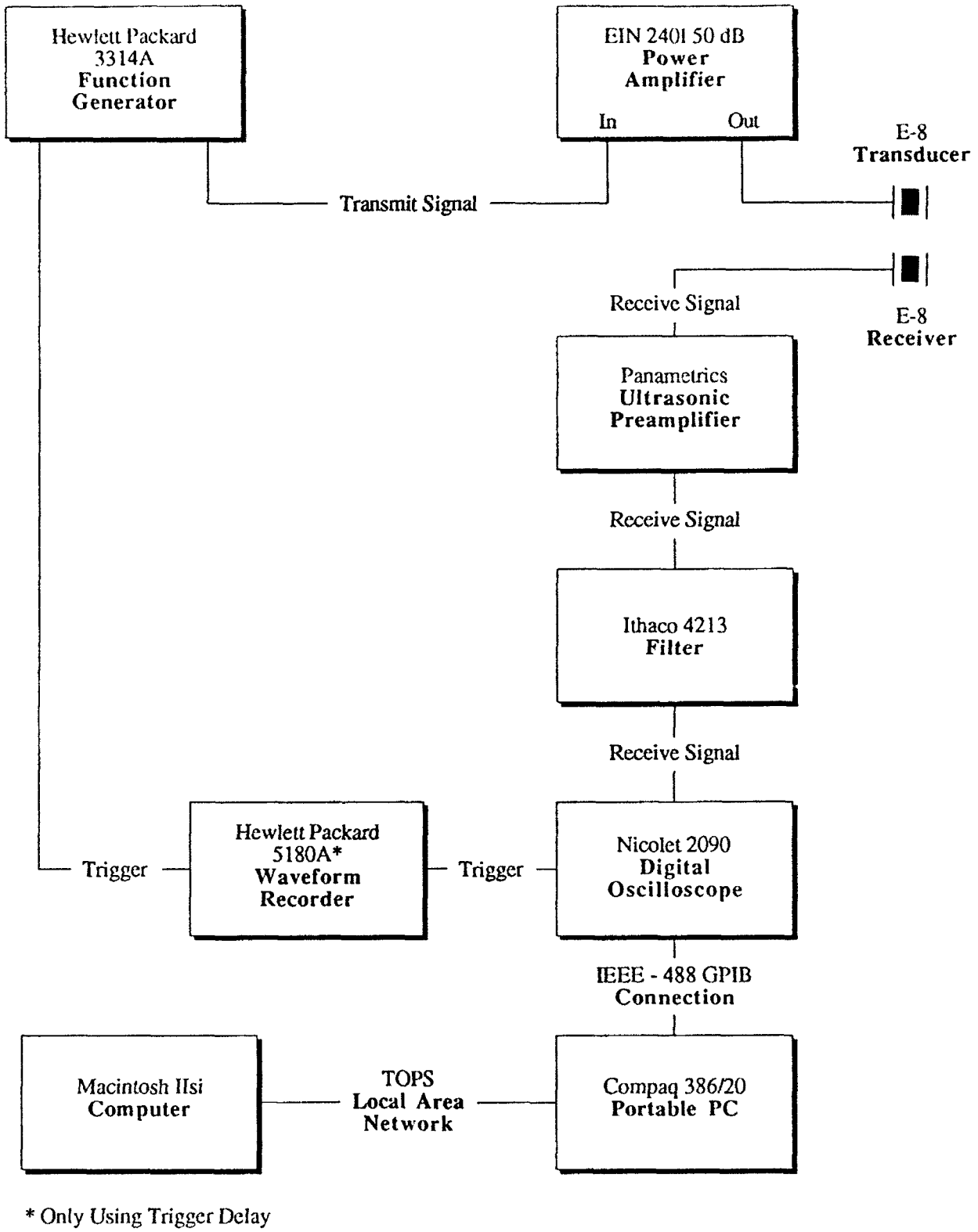


Figure 3.1-2 Tank Set-up for Bottom Reflection Signal Collection



* Only Using Trigger Delay

Figure 3.1-3 Underwater Signal Collection Hardware

the signals. Since the oscilloscope could capture 12 bits of resolution, the goal was to take advantage of its full range by increasing the amplitude of the signal as much as possible without being in danger of clipping any of its values. The gain on the preamplifier was set at either 0, 10, or 20 dB. The voltage scale on the oscilloscope could be set at 400 mV, 200 mV, 1 V or 2 V. Larger voltages meant that the incoming signal was large enough that a smaller voltage setting would produce clipping. The opposite effect existed for the preamplifier gain. Using the various hardware components' settings, the signal set was adjusted so the maximum amplitude representation possible was captured during collection.

The combinations of frequencies and sinusoid cycles used in capturing the signal sets can be seen in Table 3.1-1. The strategy was to produce signals with both a constant pulse width across the frequencies and a constant number of cycles (4) across the frequencies. Eliminating the redundant combinations, ten conditions were provided. Within each condition 16 individual signals were recorded to allow noise reduction by averaging. The signals were recorded at 2 MHz over 12 bits, with 2048 samples per signal. In addition to the Bottom and Free-field conditions, bottom-only and noise signals were recorded. The total signal set is summarized in Figure 3.1-4.

Frequency	Constant Pulse Width	Constant Pulse Width	Variable Pulse Width
200	2	(4)	4
400	4	8	(4)
600	6	12	4
800	8	16	4

where (4) = Duplicate

Table 3.1-1 Frequency and Sinusoid Cycle Combinations for Signal Collection

The oscilloscope did not provide an automatic means of adjusting the DC offset of the signal to zero, so the first step in being able to use the signals required that the DC offset be eliminated from each signal. This was accomplished by adding all points in the set of 16 instances of one type of signal and dividing each point in the signal by 16*2048. This result is then subtracted from each point in each of the 16 signals, resulting in 16 signals which are mean 0 adjusted. The adjustment was done over 16 signals because the oscilloscope was not changed between individual signal shots while the data was being collected. After removing the offset the 16 adjusted instances of each signal class were averaged to produce one averaged, mean 0 adjusted signal. The averaged signals were low-noise versions which, with further signal processing for particular needs, could

	ANGLES	SHELL THICKNESS	MATERIALS	SURFACES	FILTER SETTINGS	FREQ/CYCLE COMBINATIONS	REPETITIONS	SIGNALS
Free Field Targets	3	2	2	–	–	10	16	1920
Targets on Sand	3	2	2	–	–	10	16	1920
Smoothed Sand Without Targets	–	–	–	2	–	10	16	320
Sand With Angle Depression	3	–	–	–	–	10	16	480
Noise With Machintosh IIsi Monitor Off	–	–	–	–	3	–	16	48
Noise With Machintosh IIsi Monitor On	–	–	–	–	3	–	16	48

Figure 3.1-4 Underwater Signals Collected

be used in the human and network tasks. Any signal processing performed on the signals were based on the mean 0 adjusted signals.

3.2 AIR SIGNALS

Due to the severe initial difficulty in classifying the underwater sounds, as discussed in Sections 4 and 6, it was decided that a different approach to the data presented in the experiments could be helpful. With this in mind it was decided that the targets used in the original conditions would be used in creating a set of non-underwater returns. A sound-attenuated laboratory at the Catholic University of America was chosen as an appropriate environment for the signal set generation. The signals were created by suspending each target from a metal stand and striking it with a wand that had different materials attached to its end.

By virtue of using common targets the Air signal set shared two parameters of Material and Thickness with the Free-field and Bottom sets. Angle of insonification obviously did not apply, but was replaced by the type of Striker as the third parameter for the Air signals. The entire set consisted of striking Brass and Steel, Thick and Thin targets with either a metal, plastic, or wood instrument. Therefore, as in the two underwater cases, 12 classes of signals were created from two Materials, two Thicknesses, and three Strikers.

Unlike the highly automated collection of the Free-field and Bottom signals, the Air signals involved more manual control. Each of the four targets was hung by monofilament from two parallel horizontal arms on a vertical stand. The monofilament was shortened to reduce the amount the target could swing after being struck. A small hard-plastic wand was manufactured which could have an end-piece screwed onto it. The end-pieces were toroidal and made of either metal, hard-plastic, or wood. The signals were created by striking a hanging target with the wand fitted with an end-piece. The sounds made by striking the targets were collected using a Sennheiser 421 microphone which was attached to a Sony TCD-D10 Pro Digital Audio Tape (DAT) machine with a Shure A95U adapter.

In order to match the 16 returns collected for the Free-field and Bottom instances, many repetitions of the Air signals were generated. The process of manually striking a target and getting a noise-free return was more difficult than the automatic Free-field and Bottom collection. For each target it was empirically judged how many strikes were necessary in order to be able to get 16 final

clean signals. Typically it required between 23 and 30 strikes to ensure a good set. The signals were generated by striking the targets lightly at an angle in line with the microphone which was located below and on the opposite side of the target from the Striker.

Once all of the signals were recorded on the DAT, they had to be transferred to a Macintosh and their individual instances put into separate files. The National Instruments (NI) signal processing package, LabView, was used in conjunction with a 16-bit NI A2100 D/A data acquisition board to capture each signal class from the DAT with a sampling rate of 32 kHz. Although the signals were in an audible range, they needed some processing for consistency. The signals were extracted from the large file containing all signals in one class into separate files. During this process the signals' initial speculars were aligned and their end points were determined by a windowed thresholding process. The initial specular of a signal is the point at which the initial target return energy appears. Each of the signals was padded with 1600 points at the beginning and 16000 points at the end with points which originally separated the signals in the large class file on the DAT recording. This processing produced signals which ranged in length from 13200 and 39650 points. The extractions produced 374 separate files, each containing one Air signal.

4.0 SIGNAL PREPARATION FOR HUMAN EXPERIMENTS

An initial experiment was run soon after the collection of the underwater signals and the initial processing described in the Section 3 were completed. An appropriate set of signals was needed for the experiment and the goal was to find a set that was diverse, but that could not easily be memorized. The process required that the averaged signals at different insonifying frequencies and numbers of sinusoid cycles be evaluated.

4.1 PILOT SIGNAL PROCESSING

In order to perform the evaluation the signals had to be downsampled into the range of human hearing which is normally between the 20 Hz - 20 kHz. For the downsampling, a linear interpolation was performed at a 5:1 ratio of the original to the lengthened signals. The interpolation simply involved inserting four new points linearly between each two original points. The 10236 point interpolated signals were converted from their 12-bit original form to 16-bit amplitudes to allow the National Instrument's (NI) D/A board its maximum range. Finally, to prevent potential aliasing problems, a 600 point linear ramp was applied at both ends of each signal. The resulting signals played at a 24 kHz sampling rate were 427 ms in length, with a 25 ms ramp. The returns from the 600 kHz, 4-, 6-, and 12-cycle sinusoid insonifying pulses were chosen as a good input set. The decision was based on overall satisfaction with the relative quality of the signals in the 600 kHz set, and the fact that there were enough signals to hamper memorization.

As the main interest in the research revolved around the complication of classifying signals containing bottom reflection, the preliminary experiment was conducted using the Bottom signals. The results from the experiment revealed that this set was considerably harder to classify than anticipated and it was decided that an experiment using the Free-field signals should be run as a benchmark. The strategy behind this change lay in the assumption that due to their relatively higher signal-to-noise ratio, with no bottom reflection, the Free-field signals were innately easier to classify than the Bottom signals. The Free-field signals were evaluated, using the same processing as described for the Bottom signals, for an appropriate set of signals to use for the experiment. The 400 kHz, 4- and 8-cycle signals were chosen for two reasons. First, they were empirically the best sounding signals; and second the difficulty with the Bottom signals led us to search for a smaller, slightly less complex set.

Further examination of the signals showed that although the Free-field and Bottom signals were collected in a non-noisy environment, they had spurious frequency problems. Investigation into the matter, both by looking at and listening to the signals, revealed little scientific evidence of the cause in the case of the Bottom signals. The Free-field signals had the spurious problems that the Bottoms experienced, plus added interference from noise in the collection process. A computer monitor located two feet from the oscilloscope introduced electronic noise on the connections in the collection hardware. The monitor noise problem was discovered during the collection effort. Therefore the noise was recorded, and later analyzed so it could be extracted to the extent possible from the Free-field signals.

In preparation for removing the offending frequencies, observation of the interpolation method revealed that aliasing frequencies were being introduced during the processing. So not only did frequencies from noise and spurious sources need to be eliminated, but another method for making the signals an audible length had to be found. As the expansion of the signals was most easily addressed by interpolation, a different algorithm was determined for it that did not introduce an aliasing problem. The interpolation was to be done in the frequency domain which had the added advantage that the signals would be in the correct form to be able to have any problem frequencies extracted.

The frequency domain interpolation was performed using the following method. First an FFT was taken of a 2048 point original signal. The resulting 2048 values consisted of, in order, the dc offset, 1023 positive frequency amplitudes, the Nyquist frequency amplitude, and the 1023 negative frequency amplitudes in reverse order. An array of 16384 points was created to hold the frequency interpolated values. The dc offset was copied from the original array to the large array. The 1024 frequency amplitudes, including the Nyquist value, were copied to the large array. The Nyquist value and the last 1023 points from the original array then were copied to the last 1024 places in the large array. Finally all of the values in the large array between the original halves of the FFT frequencies were set to 0.0. Following the transfer of values an inverse FFT was performed on the large array. This processing achieved the goal of lengthening the signal without adding unwanted frequency components.

Once the frequency domain interpolation was completed, the signals needed to be scaled. The NI board's 16-bit capacity was filled by scaling each signal individually to the range (-32767, 32767).

Additionally, filtering was performed on both sets of signals to remove the offending frequencies present.

Extraction of the spurious frequencies involved performing narrow-band filtering on the frequency domain interpolated signals. Frequency spectra of the original length signals were created and examined to determine where the aberrant signal behavior was in the frequency content. Several extremely narrow-band spikes were apparent in many of the signals, and it was decided that since the spikes obviously apparently were not innately part of the signals, they could be judiciously removed individually. The process involved determining exactly how many spikes existed and in what signals for both the Free-field and Bottom reflection signal sets.

Once the frequencies to be eliminated were determined, each signal was filtered individually. The Free-field signals had both the spurious and noise induced frequencies removed while the Bottom signals needed to have only the spurious frequencies removed. The signals were filtered after the frequency domain interpolation was performed. During the filtering process it was important not to interfere with the phase of the signal, so the interpolated signals were converted from rectangular to polar coordinates, and only the magnitudes were changed. The signals were filtered below 100 Hz and above 1 MHz by setting the magnitudes for those frequency bins to 0. The magnitudes for the frequency bins affected by the spikes were altered in one of two ways. If the spike affected only one frequency bin, the magnitude was set to the average of the amplitude values of the frequency bins on either side of the affected bin. If the spike encompassed more than one frequency bin, which was a less common occurrence, a linear interpolation of the flanking bins' values was performed and the bad values were replaced with the newly interpolated amplitudes. The Free-field signals also had the monitor noise frequencies removed in the same way. The method used provided the means for eliminating any offending frequency spikes without affecting the legitimate frequency content of the target returns.

4.2 FINAL SIGNAL PROCESSING

The results from pilot experiments using the frequency domain signals described above showed that the subjects continued to have difficulty in performing the classification task. The signal set was revisited in an effort to identify factors which contributed to the difficulty. Signals from each of the three collection conditions were examined and the details are presented here.

4.2.1 Free-Field and Bottom Signal Conditions

The signal-to-noise ratio was increased to produce the cleanest possible signals for the final experiments. It was decided that the signals created with the 400 kHz 4-cycle sinusoid insonification would be used Instead of the averaged 400 kHz, 4- and 8-cycle insonified signals. The decision was made so the signals used in the experiments would be as consistent in nature as possible. However, it was necessary to avoid having a set that was so small that it would be easy to memorize the individual signals. In answer to this concern, the individual instances rather than the averaged signals were used in creating the training and testing sets. Each mean 0 adjusted individual instance from each signal class was processed in the following way to produce signals that could be used in the final experiments.

The Free-field and Bottom signal sets were treated in principally the same way, although some of the details for the two sets differed. Each original signal was 2048 points in length. A Fast Fourier Transform (FFT) was performed on the signal to convert it from the time domain to the frequency domain. The resulting FFT had a band-pass filter applied to it to eliminate the unwanted frequencies and increase the signal-to-noise ratio. The band-pass for a Free-field signal was 243.2-587.9 kHz and for a Bottom signal was 229.5-587.9 kHz. Different ranges for the filters were used due to the monitor noise present in the Free-field case which required a higher high-pass cutoff value. Once the signal was filtered an inverse FFT was applied to convert it back to the time domain.

The Free-field signals were aligned with respect to their initial specular energy to reduce the potential acoustic cue available from the location of the onset of a signal's energy. The alignment was performed automatically by searching for the point at which the amplitude of the signal exceeded 10% of its maximum. The signal was then shifted to begin 30 points prior to this excessive amplitude. Linear ramping then was used at the beginning and end of the signals to prevent aliasing that could be caused by the sudden offset or dropoff of energy. The 30 point shift provided enough points to apply an increasing linear ramp to the first 25 points while the last 5 points ensured that any minor portion of the specular was included, but not ramped. The end of the signal had a decreasing linear ramp applied to it as well. The end ramp was started at different points for the signal classes, depending on where the energy for the signal fell to noise levels. The classes and the points where the ramp was started are listed in Table 4.2.1-1. The ramp continued past the points listed in the table for a total of 100 ramped points in each signal.

Signal Class	Starting Point of End Ramp
B10	550
B14	700
B19	500
B50	550
B54	700
B59	500
S10	500
S14	700
S19	500
S50	600
S54	700
S59	500

Table 4.2.1-1 Signal Classes and their Initial Ramping Points

The Bottom reflection signals did not require that an alignment be performed. The first 25 points of the signals were increasingly linearly ramped, again to avoid any potential aliasing problems. The signals also were decreasingly ramped in the same manner as the Free-field set. Here the linear ramp started at point 1730 in each of the signals, and continued for a total of 100 points.

The remaining processing was identical for both sets of signals. The aliasing problem discussed earlier caused by linear interpolation of a signal was resolved by performing what could be referred to as a frequency domain interpolation. The principle here was to increase the resolution of the signals without altering their frequency spectra. To do this, an FFT was taken of a 2048 point signal. The resulting 2048 values were the typical output from an FFT routine. They consisted of, in order, the dc offset, 1023 positive frequency amplitudes, the Nyquist frequency amplitude, and the 1023 negative frequency amplitudes in reverse order. An array of 32768 points was created to hold the frequency interpolated values. The dc offset was copied from the original array to the large array. The 1024 frequency amplitudes, including the nyquist value, were copied to the large array. The nyquist value and the last 1023 points from the original array then were copied to the last 1024 places in the large array. Finally all of the values in the large array between the original halves of the FFT frequencies were set to 0.0. Following the transfer of values an inverse FFT was performed on the large array. This processing achieved the goal of increasing the number of points in the signal without adding unwanted frequency components. Once the frequency domain interpolation was completed, the only remaining issue was scaling. To take full advantage of the range of the NI board's 16-bit capacity, each signal was scaled individually to the range (-32767,

32767). The resulting signals were then in good condition to be used in the psychoacoustic experiments.

4.2.2 Air Signal Condition

The description of the collection of the Air signal set in Section 3 reveals that the Air signals required relatively little processing in order to prepare them for use in the experiments. The signals, audible to humans by default, were sampled at 32 kHz and could be played at 32 kHz over the A/D board, so no sampling changes were needed. They were also already a suitable length for human subjects, so the duration of the signals did not need alteration. Custom software written with the D/A board's LabDriver library of functions was used to listen to each return in a signal class to determine a set of 16 clean, consistent signals to use for each class in the experiment. The signals were chosen based on the clarity and quality of the return. Since the insonification of the targets was not automatic, it was important not to include any signal which contained artifacts that were not part of the return energy. A set of 96 signals was selected, 12 classes by 16 instances, where half was used for the training set and half for the testing set for the experiments. The hardware setup used for listening to the signals was identical to that used in the psychoacoustic experiments and is described in Section 5.

5.0 HUMAN ACOUSTIC CLASSIFICATION EXPERIMENTS

The acoustic signals described in the previous sections were used in psychoacoustic experiments which were conducted on one sonar-experienced and one novice set of human subjects. The sessions of the experiments were run in a laboratory setting over the course of several weeks. There were three conditions for the experiments, one for each of the Free-field, Bottom reflection, and Air signal sets. The experiment was conducted in the same manner for all conditions and for both subject groups, with only the data being changed. Each condition required that subjects participate in seven training sessions and one test session.

5.1 CLASSIFICATION TASK

The experiment task involved listening to and classifying a set of signal returns. The three parameters to be classified for each target were Material, Thickness, and either Angle or Striker, depending on whether the signals were from the underwater or air environment respectively. As described earlier in the Section 3, the target material was steel or brass, and the shell thickness was either "Thin" (5% of the exterior diameter of the shell), or "Thick" (10% of exterior diameter). The Free-field and Bottom targets were insonified at three angles with respect to the beam of the pulse: 90° (broadside), 45°, and 0° (along the axis of the target). However, Angle did not apply in the case of the Air signals. These targets were excited by strikers with tips made of metal, plastic, and wood. Each of the three parameters was identified for all signals presented in the experiment.

5.2 HARDWARE

The experiment required a variety of hardware components. The instruction screens were shown and the subjects' responses saved on a Macintosh IIsi computer. The signals were played using a National Instruments (NI) A2100 A/D board located in the IIsi. The NI board was attached to an NAD 7225PE receiver used for amplification and volume control. The subjects then heard the sounds through Sennheiser HD 250 linear headphones.

5.3 INPUT DATA

The signals used in the Free-field and Bottom conditions of the final experiments were the 400 kHz, 4-cycle sinusoid returns. They were played for the subjects at a sampling rate of 16 kHz.

The Air condition experiments used returns from the Air signal set played at 32 kHz. The difference in the sampling playback rates stemmed from the innate difference between the signal conditions. The Free-field and Bottom signals were played at the slowest rate on the A/D board to expand them as much as possible. This rate was judged empirically to provide the most opportunity to gain information from the signals. The Air signals' original capture sampling rate was 32 kHz, so that was what was used for playing these signals for the subjects.

Signals for the three conditions were divided into training and testing sets, each made up of eight of the individual instances for each of the 12 signal classes. The training set of instances 1-8 was used for each of the seven training sessions, while the testing set of instances 9-16 was reserved for the test session. Three instances from each class in the training set were chosen randomly for each of the training sessions for each subject. In addition, each training session had a different randomization of 36, of a possible 96, signals presented. During the test session, however, all 96 signals from the testing set were randomly presented.

5.4 SESSIONS

The first session included an orientation portion that was not included in the remaining sessions. First this involved the subject's acclimation to the manner in which the experiment interface worked. Second, and more importantly, the subject was presented with a random sample of the 36 of the signals used in the sessions, where three signals were from each class. During this presentation the subject was not required to make any classification judgments. After the orientation the subject went on to the main task of listening to and classifying the parameters of each of the signals presented. The second through seventh training sessions and the test session included only the main portion of the first session where the signals were actually classified. The classification process itself is described next.

The experiment sessions were presented on a Macintosh IIsi with a graphical user interface for the instruction screens. An A/D converter board, a stereo receiver and headphones, all described above, were used for playing the signals. The subjects read the screen for instructions and used the mouse to make selections via buttons on the screen. Examples of the screens are shown in Figure 5.4-1. The subjects were allowed to adjust the volume and balance, but no other controls, on the receiver at any time during the sessions.

Figure 1(a)

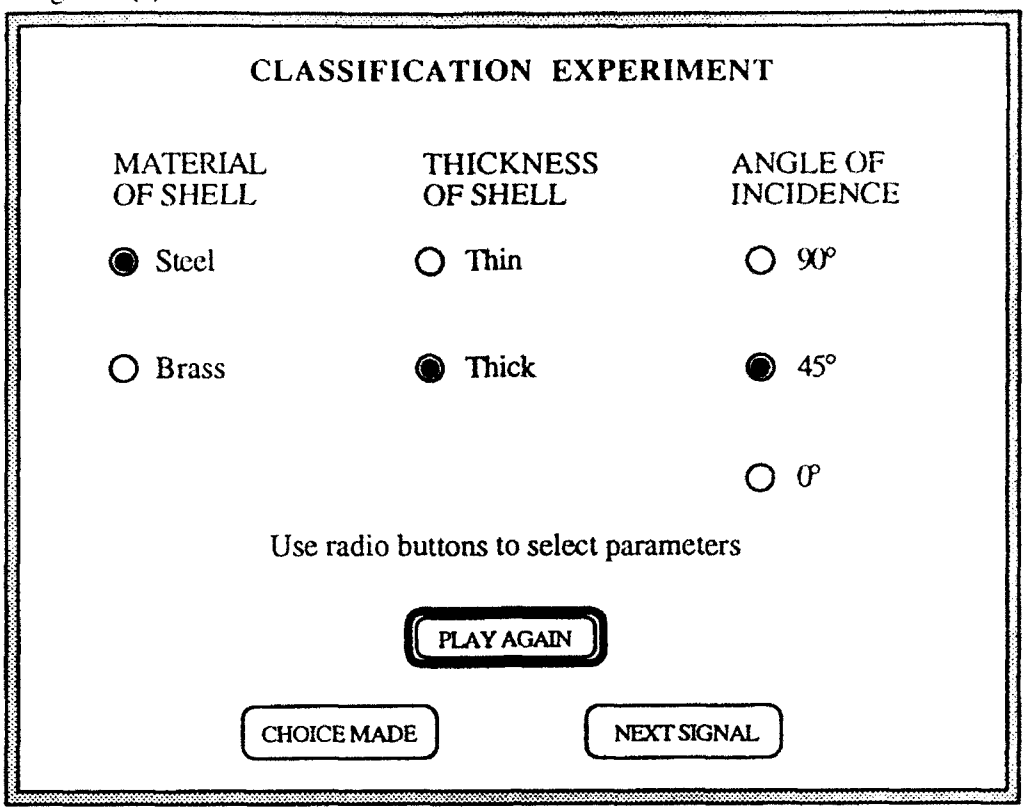


Figure 1(b)

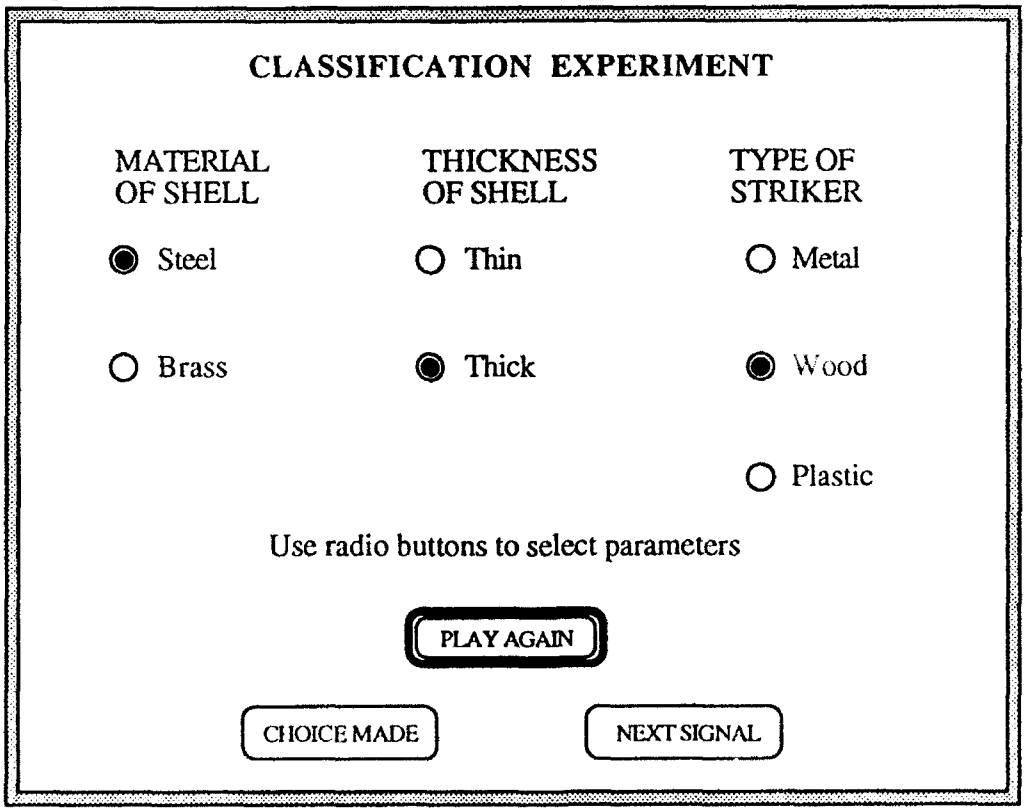


Figure 5.4-1 Classification Experiment Screens

In the classification portion of the sessions the subject could listen to each signal as many times as desired. To guarantee that the signal was heard at least once, it was played automatically before any parameter choice was allowed. After making selections for each of the three parameters the subject clicked a button to continue to the next signal. At this point in the training sessions the subject received feedback as to the correct parameters for the current signal, and heard that signal played again. The signal could be played even more times at this point, or the subject could choose to go to the next signal. In the test session, however, the subject's choice to continue brought up the next signal without feedback or hearing the current signal. Feedback was assumed to promote further learning, so it was eliminated from the test sessions. The purpose for the difference was to test the subject's knowledge of the characteristics learned about the signals during the training sessions.

During all sessions the subject's responses for the parameters were recorded and stored. The data stored for each subject for each session included the randomization order of the signals, the subject's responses to the individual parameters for each signal, and whether the subject was correct on all three parameters simultaneously. The data in these files were used in the analysis of the human's classification performance and strategies detailed in the following sections.

6.0 RESULTS OF PSYCHOACOUSTIC EXPERIMENTS

The performance data from two subject groups and three signal conditions are presented here. Subjects with and without sonar experience were tested, to see if that experience was correlated with any performance differences in the classification task.

6.1 EXPERIMENT SUBJECTS

As mentioned earlier, two sets of subjects, one with and one without sonar experience, participated in the acoustic experiments. The subjects with experience were sonar technicians from the United States Navy who were recruited by Dr. David Kobus from the Naval Health Research Center (NHRC) in San Diego, California. They ranged in age from 24 to 39 and their sonar experience varied from 3.5 to 14 years. Ten sonar technicians participated in the experiment where each subject ran the eight sessions, seven training and one test, for each of the three signal conditions. The subjects were randomly assigned an order of conditions from a counterbalanced schedule. The purpose was to minimize any possible order effect that might occur in the subjects' performance. The condition order and two personal statistics for the group are shown in Table 6.1-1. Although the conditions were counterbalanced for the group, any order effect that may have occurred did not adversely affect the results since comparing the performance for the two groups was not a main goal of the study.

Subject	Week 1	Week 2	Week 3	Age	Years Sonar Experience
1	B	A	F	39	3.5
2	A	F	B	33	8
3	A	B	F	33	7
4	B	F	A	26	7
5	B	F	A	34	12
6	A	B	F	33	7
7	F	B	A	32	14
8	A	F	B	24	5.5
9	F	A	B	NA	7
10	F	B	A	NA	6

Table 6.1-1 Counterbalanced Condition Randomization and Experience for Navy Subjects

For each signal condition, all eight sessions were run in one week. Generally, one to two training sessions were run per day. On the last day at least one training session was scheduled, followed

by the test session. This guaranteed that the subject's memory of the signals was refreshed before the test session was executed. The two remaining conditions then were run in subsequent weeks.

The inexperienced subjects were students at the Catholic University of America in Washington, D.C. who ranged in age from 18 to 22. The students were run as a pilot group so each student ran the eight sessions for only one signal condition of the experiment. There was a total of four student subjects per condition, with 12 students completing the sessions. Since the students did not participate in a counterbalanced randomization of all conditions, the condition for each subject was chosen based on the primary goal of getting four subjects to complete the experiment for each condition. The sessions for the students were scheduled in the same way as for the experienced subjects with all sessions being performed within one week's time. Again, at least one training session was administered on the last day prior to the test session.

6.2 SUBJECT PERFORMANCE

Subject performance varied considerably across the three conditions, as expected from the pilot experiments. Performances are considered statistically above chance at the 5% level if they exceed the values given in Table 6.2-1. These figures are derived from a grouped t-test.

	Material	Thickness	Angle/Striker	Overall
Chance	50	50	33	8.33
Training Session	67	67	52.77	25
Test Session	61	61	43.75	16.67

Table 6.2-1 Chance and Statistically Above Chance Percentages for Different Experiment Sessions

6.2.1 Free-Field Results

The performances of the subjects on the final test session of the Free-field experiment are shown in Table 6.2.1-1. These data are graphed for the student subjects in Figure 6.2.1-1 and for the Navy subjects in Figure 6.2.1-2.

There is little evidence that any subjects were able to distinguish Material or Thickness. Only one subject in each group had a Thickness test score significantly above chance, and there were no Material test scores above chance. Thirteen of fifteen subjects were able to discriminate Angle at levels significantly above chance. Casual listening suggests that it is easiest in the Free-field data

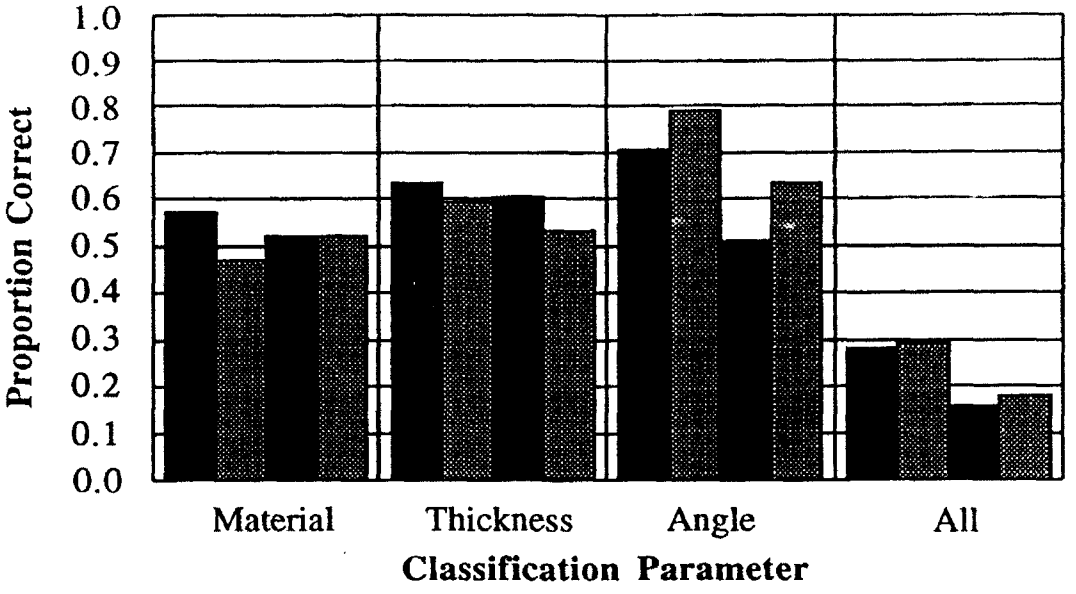


Figure 6.2.1-1 Free-Field Test Session Performance for Student Subjects

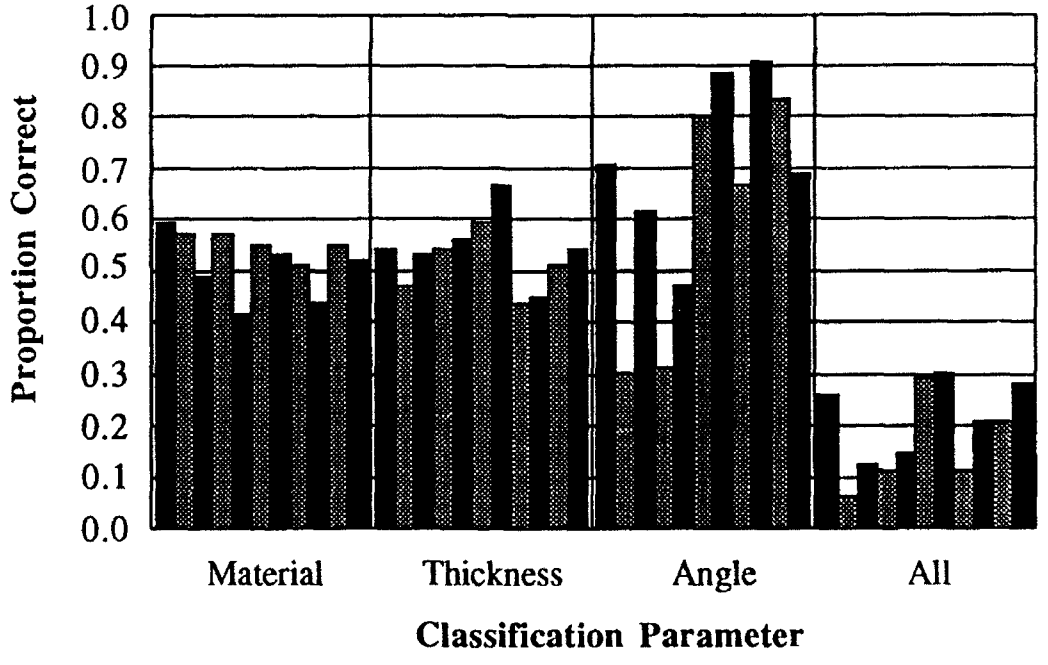


Figure 6.2.1-2 Free-Field Test Session Performance for Navy Subjects

to identify 90° signals, due to their short duration. A subject who had learned to discriminate the 90° signals from others, but could not tell 45° signals from 0° signals, would be expected to have near a 0.67 performance level. Navy subjects 3, N3, and 4, N4, learned to discriminate many of the 45° and 0° signals as well, since their scores are both near 0.90 correct. Two other Navy subjects are also above the 0.67 level, indicating some knowledge of the 0° and 45° signals. The bulk of the subjects, however, were unable to learn more than the characteristic of the 90° signals. In several cases the higher performance on Angle was enough to make the overall classification performance statistically higher than chance. Figure 6.2.1-3 shows the Navy subjects' performances by session, averaged across all subjects.

CATHOLIC

Subject	Material	Thickness	Angle	Overall
7	0.57	0.64	0.71	0.28
9	0.47	0.59	0.79	0.29
11	0.52	0.60	0.51	0.16
19	0.52	0.53	0.64	0.18
Mean	0.52	0.59	0.66	0.23
Std Dev	0.04	0.04	0.12	0.07

NAVY

Subject	Material	Thickness	Angle	Overall
7	0.59	0.54	0.71	0.26
11	0.57	0.47	0.30	0.06
10	0.49	0.53	0.61	0.13
9	0.57	0.54	0.31	0.11
2	0.42	0.56	0.47	0.15
8	0.55	0.59	0.80	0.29
4	0.53	0.67	0.89	0.30
5	0.51	0.44	0.67	0.11
3	0.44	0.45	0.91	0.21
1	0.55	0.51	0.83	0.21
6	0.52	0.54	0.69	0.28
Mean	0.52	0.53	0.65	0.19
Std Dev	0.06	0.07	0.21	0.08

Table 6.2.1-1 Free-Field Test Session Performance for Both Groups of Subjects

These curves indicate that little learning took place after the first session. When the three parameters are considered separately for the first seven sessions for both sets of subjects, the

scores for the first session are significantly lower than for subsequent sessions. This is the only significant learning effect.

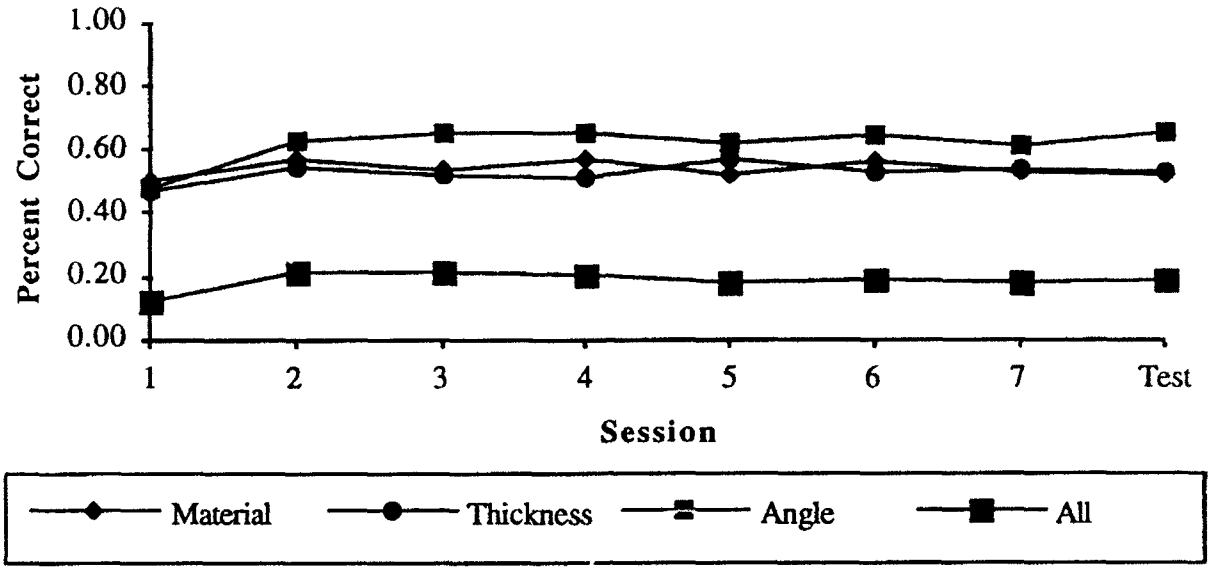


Figure 6.2.1-3 Free-Field Performance by Session, Averaged for All Navy Subjects

Students do not perform statistically differently than Navy subjects on the Free-field test. This applies to all three parameters individually as well as overall scores. Angle scores are significantly higher, as expected from casual listening.

6.2.2 Bottom Results

The performances of the subjects on the final test session of the Bottom experiment are shown in Table 6.2.2-1. These data are graphed for the student subjects in Figure 6.2.2-1 and for the Navy subjects in Figure 6.2.2-2.

The Bottom experiment also proved quite difficult. Of all 14 subjects, only one had a test score significantly above chance on Thickness, while two scored significantly above chance on Material. As with Free-field signals, these parameters are very difficult to distinguish.

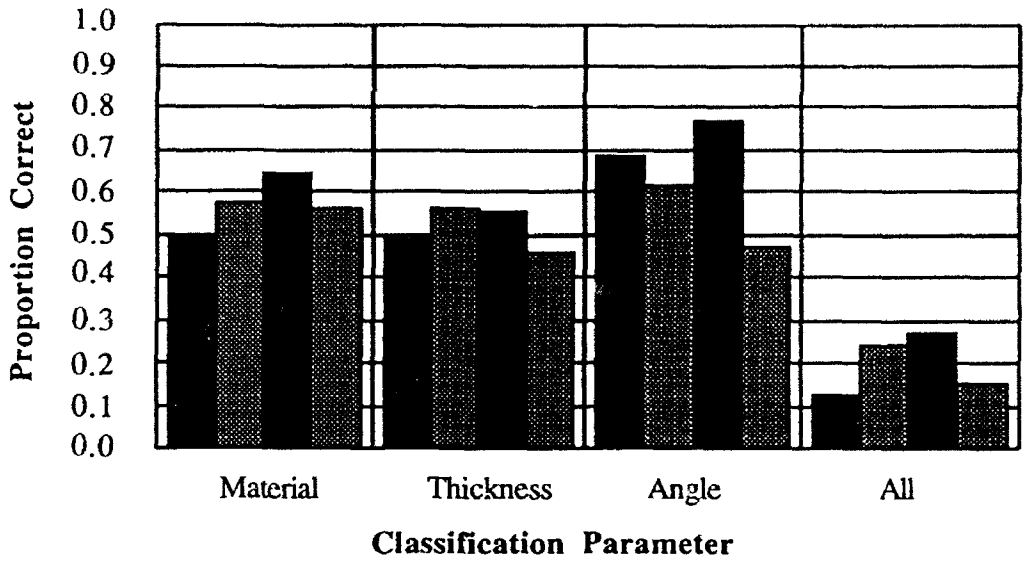


Figure 6.2.2 -1 Bottom Test Session Performance for Student Subjects

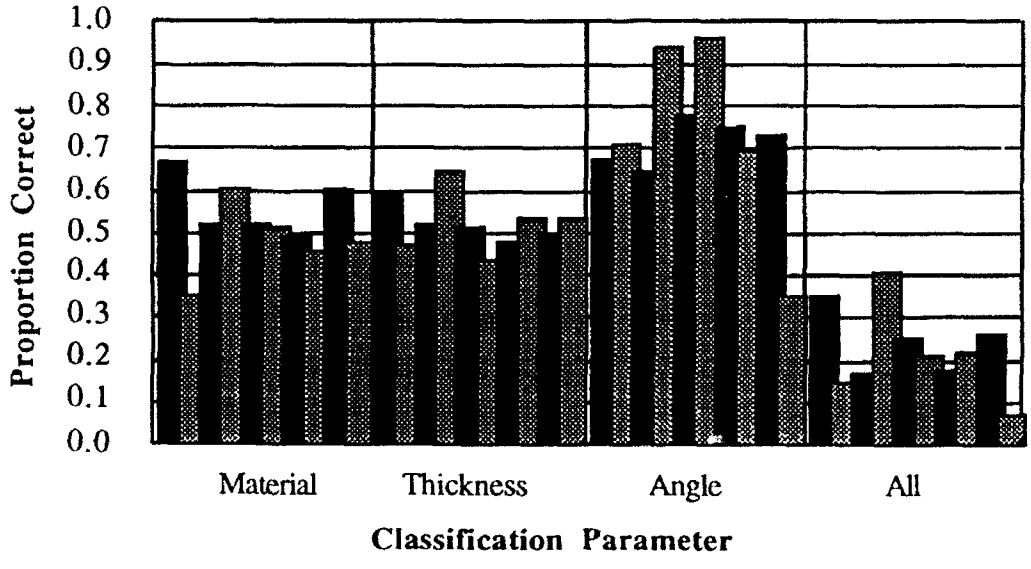


Figure 6.2.2-2 Bottom Test Session Performance for Navy Subjects

CATHOLIC

	Subject	Material	Thickness	Angle	Overall
	7	0.49	0.50	0.69	0.13
	9	0.57	0.56	0.61	0.24
	11	0.65	0.55	0.77	0.27
	20	0.56	0.46	0.47	0.16
Mean		0.57	0.52	0.64	0.20
Std Dev		0.06	0.05	0.13	0.07

NAVY

	Subject	Material	Thickness	Angle	Overall
	4	0.67	0.59	0.68	0.35
	5	0.35	0.47	0.71	0.15
	1	0.52	0.52	0.65	0.17
	6	0.60	0.65	0.94	0.41
	7	0.52	0.51	0.78	0.25
	3	0.51	0.44	0.96	0.21
	10	0.49	0.48	0.75	0.18
	2	0.46	0.53	0.70	0.22
	8	0.60	0.50	0.73	0.26
	9	0.48	0.53	0.35	0.07
Mean		0.52	0.52	0.72	0.23
Std Dev		0.09	0.06	0.17	0.10

Table 6.2.2-1 Bottom Test Session Performance for Both Groups of Subjects

Thirteen of fourteen subjects discriminated the Angle of the Bottom signals at levels above chance. As with the Free-field signals, 90° signals are relatively easy to identify. They contain a transient which stands out from the bottom reflection to the casual listener. If a subject could only tell 90° signals from the other angles, 0.67 performance would be expected. Two of the Navy subjects performed very highly on Angle, at levels of 0.94 and 0.96. Clearly these two subjects could tell 0° signals from 45° signals as well as identifying the 90° signals.

Eight Navy subjects and two student subjects scored significantly higher than chance during the test session for the parameters overall, i.e. as a simultaneous group. This performance is attributable to the high performances on Angle. The Navy subjects' average performance across the sessions is shown in Figure 6.2.2-3. The high performance on Angle is apparent, and Angle is the only parameter that shows an increase in performance across the sessions.

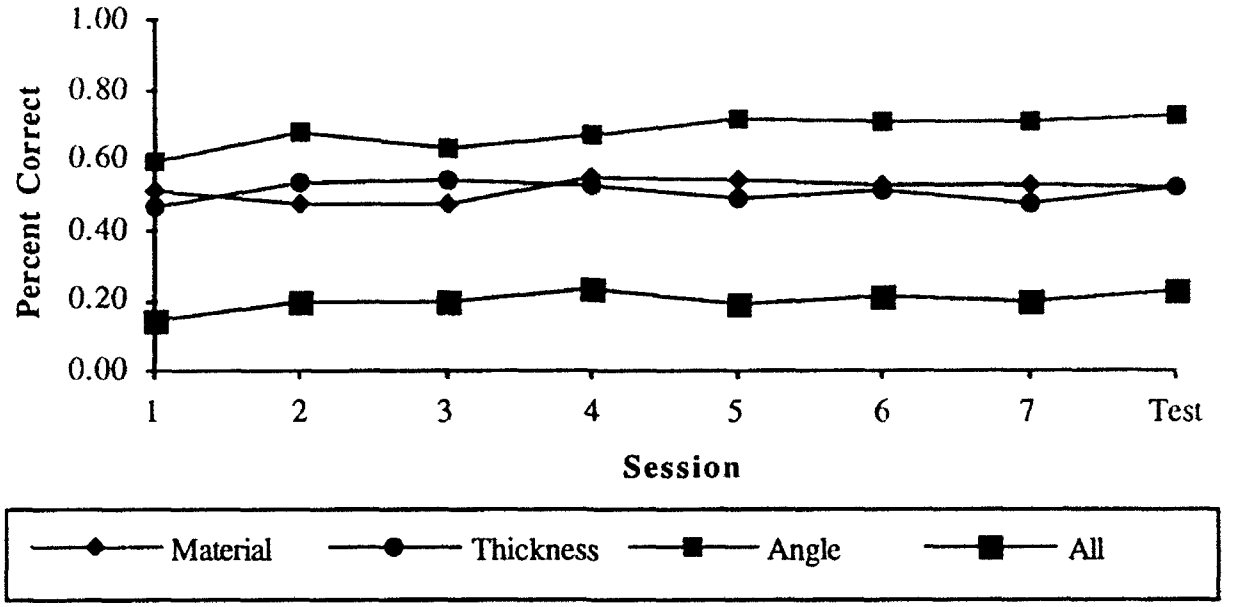


Figure 6.2.2-3 Bottom Performance by Session, Averaged for All Navy Subjects

Analysis of normalized data from the test sessions shows no significant differences between the two groups of subjects on any individual parameter for the Bottom signal condition. When the subjects' performance on the three parameters is considered over the seven training sessions, Navy performance is not significantly higher than student performance. The higher performance of the Navy group on Angle cannot be considered significant at p=0.0654. Significant learning effects are noted between session one and sessions four, five, and six when all subjects are considered.

Although the Navy subjects do not perform significantly higher than the students when the individual parameters are considered over the training sessions, when the 'Overall' performance is considered the Navy subjects did perform significantly higher. The Angle parameter, although not significantly higher for Navy subjects than students, is the only contributing factor to the significantly higher performance Overall. This difference is apparently due to the ability of two Navy subjects to discriminate 0° and 45° signals as well as 90° signals.

6.2.3 Air Results

Performance results for both subject groups on the Air signals are shown in Table 6.2.3-1, and graphed in Figures 6.2.3-1 and 6.2.3-2.

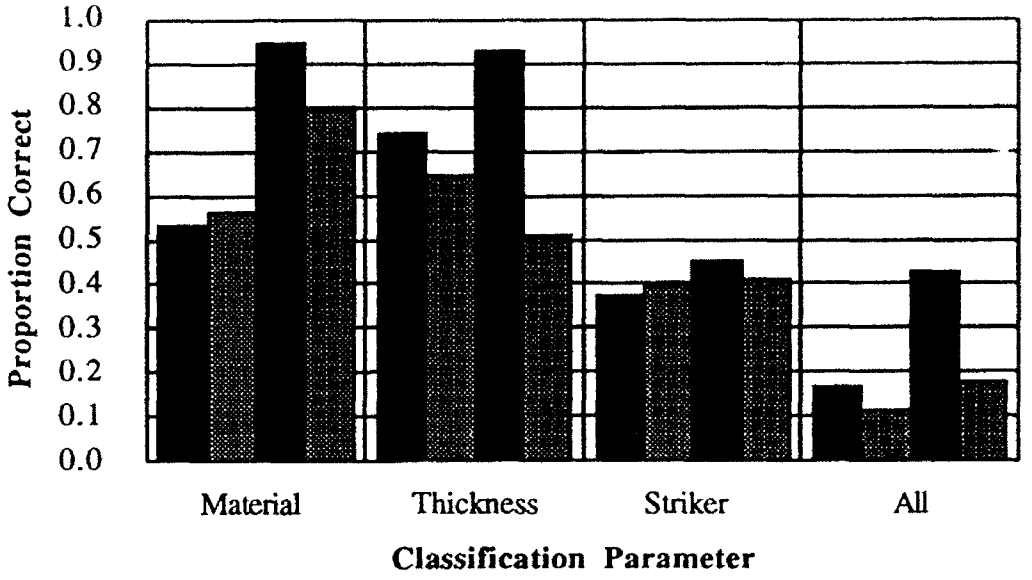


Figure 6.2.3-1 Air Test Session Performance for Student Subjects

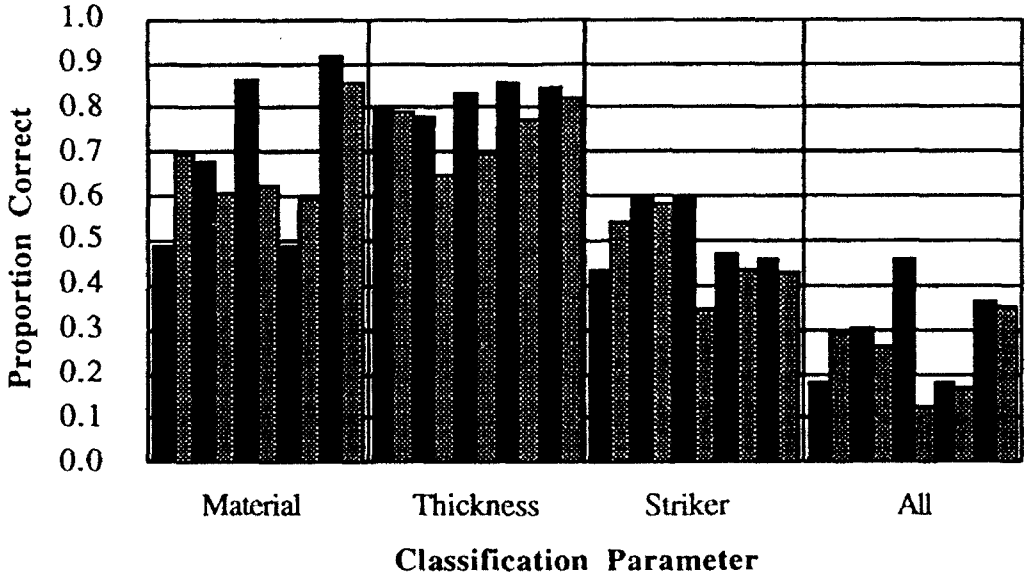


Figure 6.2.3-2 Air Test Session Performance for Navy Subjects

CATHOLIC

Subject	Material	Thickness	Angle	Overall
12	0.53	0.74	0.38	0.17
14	0.56	0.65	0.40	0.11
15	0.95	0.93	0.45	0.43
17	0.79	0.51	0.41	0.18
Mean	0.71	0.71	0.41	0.22
Std Dev	0.17	0.15	0.03	0.12

NAVY

Subject	Material	Thickness	Angle	Overall
2	0.49	0.80	0.44	0.18
8	0.70	0.79	0.54	0.29
6	0.68	0.78	0.59	0.30
3	0.60	0.65	0.58	0.26
4	0.86	0.83	0.59	0.46
1	0.63	0.70	0.34	0.13
9	0.49	0.85	0.47	0.18
5	0.59	0.77	0.44	0.17
7	0.92	0.84	0.46	0.36
10	0.85	0.82	0.43	0.35
Mean	0.68	0.78	0.49	0.27
Std Dev	0.15	0.07	0.09	0.11

Table 6.2.3-1 Air Test Session Performance for Both Groups of Subjects

Performance on the Air signals is relatively high compared to performances on the underwater signals. Unlike in the underwater condition, subjects found Material and Thickness relatively easy to discriminate. Two of four students performed significantly higher than chance on Material during the test session, as did six of ten Navy subjects. Three students were higher than chance on Thickness, as were all ten Navy subjects. One student performed higher than chance on Striker, while eight Navy subjects did so. Three students and nine Navy subjects were correct on all parameters (Overall) in the test session more often than chance performance would indicate. Figure 6.2.3-3 shows the average performances of the Navy subjects over the course of the sessions.

The high performances on Material and Thickness stood out. Also there was an apparent learning effect over the sessions, with a substantial increase in performance at the fifth session. Eight of the ten Navy subjects increased their performances from the fourth to the fifth sessions.

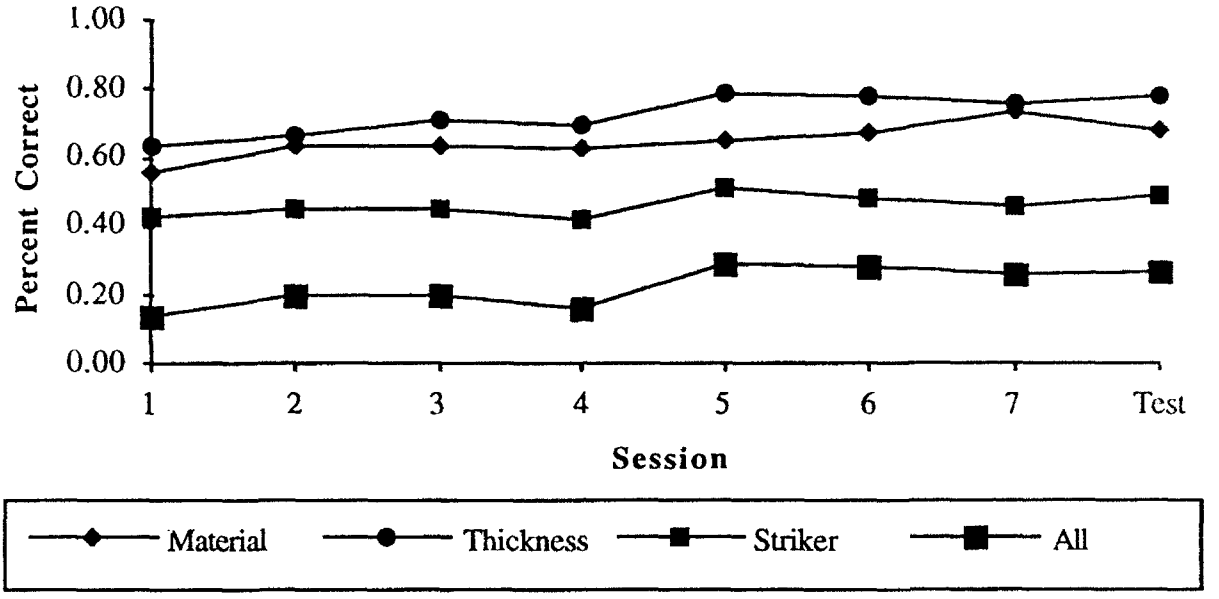


Figure 6.2.3-3 Air Performance by Session, Averaged for All Navy Subjects

When the normalized data are analyzed for differences in performance, the test sessions show no significant differences on any parameter between student and Navy subjects. It would appear that the quantity of data from the test sessions is insufficient to overcome the variability of the data, and find the higher performances of Navy subjects on Thickness and Striker significant. Nor are the Navy subjects significantly higher when considering data from all three parameters simultaneously.

6.2.4 Comparison of Navy and Student Subjects

A different picture emerges when we considered the training sessions rather than the test sessions. Considering only training sessions we examined the data for effects of subject group (Navy or student), session (excluding the test), and parameter. The Navy subjects performed significantly higher than the student subjects when considering all parameters simultaneously. Breaking this difference down by parameter, Thickness and Striker appear to be the contributing parameters, as shown in Figure 6.2.4-1.

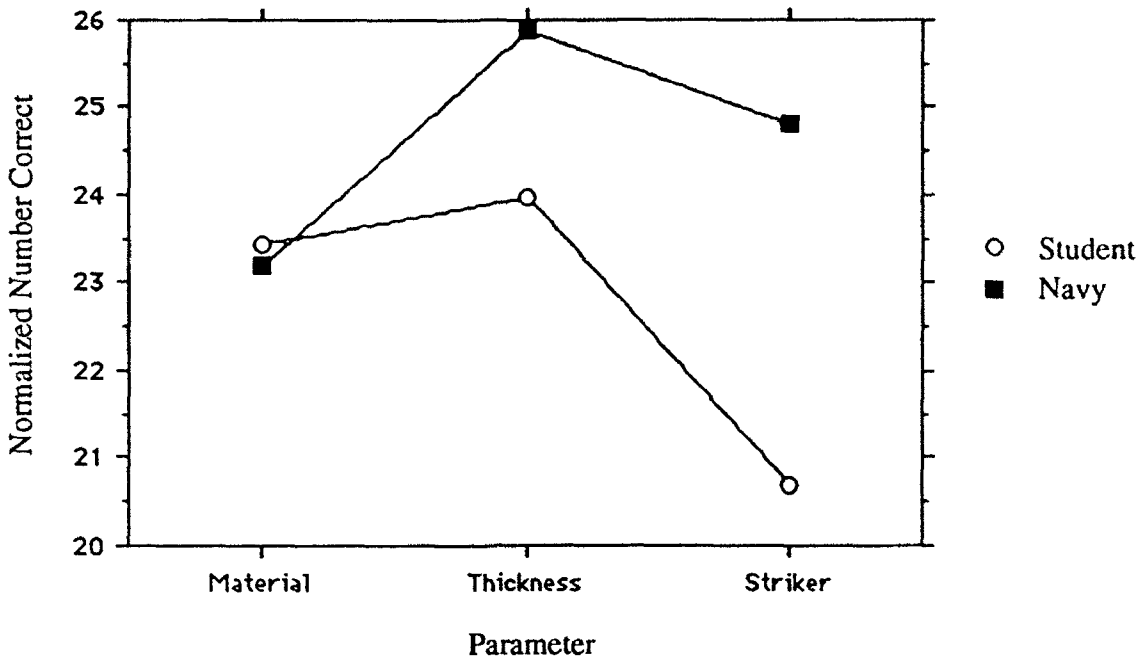


Figure 6.2.4-1 Average Navy vs. Student Subjects' Performance By Parameter

There is no significant difference between the performance of the two subject groups on Material. The difference on Thickness is also not significant (p=0.0577). The Striker difference however is quite significant (p=0.0001) with the Navy subjects higher.

There is also a significant learning effect between certain sessions. There are significant increases in performance between the sessions in Table 6.2.4-1.

Session	Higher Performance Sessions
1	3, 5, 6, 7
2	5, 6, 7
4	5, 6, 7

Table 6.2.4-1 Performance Increase Across Sessions Per Individual Parameter

Finally we examine the data concerning performance on all parameters simultaneously, that is, getting all three parameters correct ("Overall"). Here, again, we see a significant difference between the two subject groups with the Navy group performing higher than the student group. That is, the Navy subjects more often correctly identified all three parameters simultaneously than did the students. There were also significant differences between the performances on certain sessions. These data are shown in Table 6.2.4-2.

Session	Higher Performance Sessions
1	5, 6, 7
2	5, 6, 7
3	5, 6
4	5, 6, 7

Table 6.2.4-2 Overall Performance Increase Across Sessions

A plot of performances by subject group and session illustrates these differences, as seen in Figure 6.2.4-2.

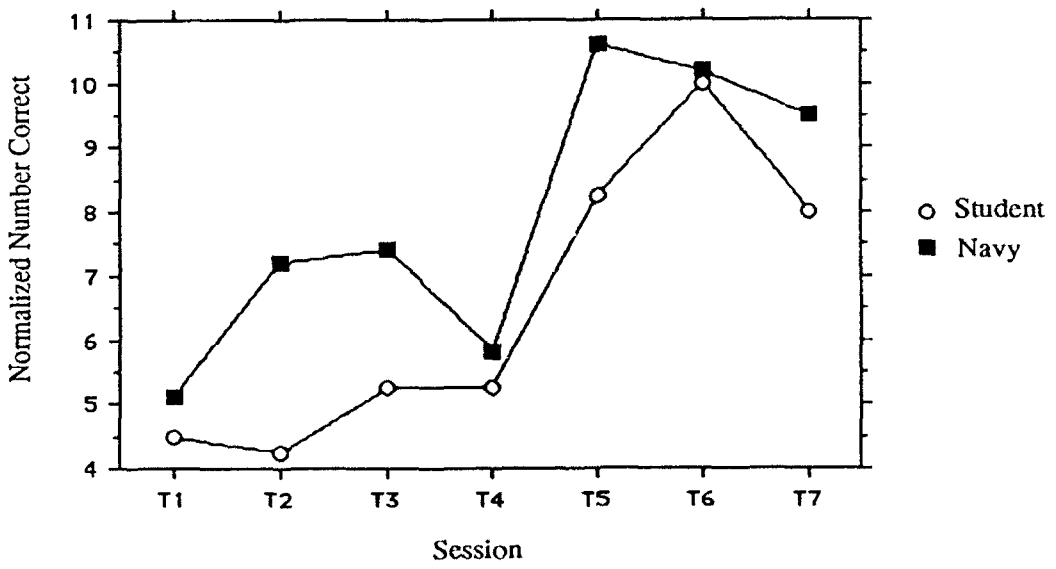


Figure 6.2.4-2 Average Navy vs. Student Subjects' Performance By Session

The differences between the Navy and student subjects emerged as the aural discrimination task became less difficult. The Free-field and Bottom tasks were extremely difficult, affording little information on which to make any discrimination. What information was present in those signals was relatively obvious to most listeners, and was detected by both subject groups. Nevertheless, two Navy subjects were able to extract enough information from the Bottom signals to discriminate between 0° and 45° signals. This is a task it is reasonable to assume no other subjects were able to perform. When the easier task of classifying Air signals is presented, differences between the two populations emerge. The Navy subjects are presumably better at extracting the information present in these signals, as long as there is enough information with which to work.

6.3 DISCUSSION

The performance results corroborate earlier pilot results as well as the impressions of the casual listener that the underwater signal classes are very difficult to distinguish from one another. The difficulty of the tasks suppressed most potential differences between the subject groups, although the Navy group performed significantly higher over the training sessions of the Bottom condition when all parameters were considered simultaneously.

The Air signal classes proved more distinct to the subjects, as the performance figures indicate. At this difficulty level more performance differences between the subject groups are significant. When considering the training session data, Navy subjects had higher performance than student subjects on the Striker parameter. The difference on the Thickness parameter was almost significant, while performances on Material were almost the same. When the three individual parameters are considered as a group, Navy subjects performed significantly higher. Navy subjects also performed better than the student group in correctly classifying the three parameters simultaneously.

7.0 SCALING

The results from the psychoacoustic experiments were analyzed using the ALSCAL multidimensional scaling model. Multidimensional scaling (MDS) is a statistical technique for discovering the pattern or structure contained implicitly in a set of data, and for representing this structure in a geometrical form. ALSCAL uses an alternating least squares procedure to determine the configuration of objects in multidimensional space which minimizes a goodness-of-fit measure. In the case of this research the "objects" were acoustic signals and the data presented to the MDS algorithm were the confusion matrices containing the subjects' judgments of the signal parameters. Complete descriptions of the MDS algorithms can be found in Young and Harris[1] and Young and Hamer[2].

Multidimensional scaling was used as an analysis tool for deriving features of the signals from the human judgment data. Scaling produced dimensions which reflected the similarities and differences found in the subjects' confusions when classifying the signals. Observation of the distribution of the signals on the dimensions provided insight about the signals and which parameters were easier or harder for subjects to identify. Signals that were similar, in the perception of the subjects, were found in close proximity to one another while the opposite was true for dissimilar signals. Each dimension revealed different ways in which the signals were grouped, and presumably different features of the signals. Combinations of the placement of signals on the separate dimensions could be used to discern the features important in classifying the signals and their separate parameters of Material, Thickness, and Angle/Striker. These issues are explored in the remainder of the session as the scaling methods and solutions are detailed.

7.1 SELECTION OF SUBJECTS AND SESSIONS

Of the ten NHRC subjects who completed all sessions for each of the three signal conditions, three were chosen as the best performers for each condition. The test session results for three "Best" subjects, chosen on the basis of their test session performance as well as on their high performance for the parameters of greatest interest in the subsequent analyses, were used as input for one set of scaling runs. Subjects 4, 6, and 8 were used for the Free-field scaling runs; subjects 3, 4, and 6 for the Bottom; and subjects 4, 7, and 10 for the Air. Their test session performance levels, and chance levels for the test sessions were seen in the Tables and Figures throughout Section 6. Another set of runs was performed for each condition for the single subject who had the highest

overall session performance. For the Free-field and Air cases subject 4 was the top performer; but for the Bottom condition subject 6 was best. Three training sessions and the test session were used in each of the scaling runs for the single top performers. The sessions included in the runs, the performance levels for those sessions, and the chance levels for the training and test sessions are listed in Table 7.1-1. The following is a discussion of creation of the scaling solutions based on this data, the solutions' dimensions and the signals' distribution along them, and the subject weights and their implications from the individual differences scaling model run on the data. For simplicity, the NHRC subjects included in these runs will be referred to as Nx where x is the subject number.

7.2 SCALING INPUT

During sessions of the experiment subjects made judgments as to the Material, Thickness, and Angle or Striker parameters for each signal presented. These responses were used as the basis for the input data to the scaling algorithms. The data were tallied in a way in which they could be viewed as similarity measures of the signals. In other words, each instance of a signal being confused with a different signal (i.e., an incorrect classification) contributed to the summation of the number of confusions of those two signals, and thus the two were assumed to be similar to each other. Since the scaling algorithms give more stable solutions using matrices of dissimilarity ratings, the data were converted into dissimilarities to be used as input.

To create a matrix of dissimilarity data the similarity ratings for each session first were collapsed into matrix form. Each matrix was 12x12 where the rows represented the actual signal classes and the columns the judged signal classes. For instance, if a subject heard an instance of a Brass 10% 90° (B19) signal and identified it as a Brass 5% 90° (B59) signal, then the B19 row, B59 column was incremented by one. After all of the signals for one session were tallied, the matrix contained the ways in which the signals were confused by the subjects. The similarity ratings in the matrices then were converted to dissimilarity ratings. The conversion was performed by subtracting each element in the matrix from the maximum total possible per element. In the case of the training sessions, the maximum was three because three instances of each signal class were presented. In the same vein, eight was the maximum possible for each test session. Each matrix filled with dissimilarity data was folded to make a lower triangular matrix that was used as input to the scaling algorithms. An example of the input matrix created from the test session results from N4 for the Air signal condition is shown in Table 7.2–1.

BEST FREE-FIELD

Subject	Material	Thickness	Angle	All
N4	53.13	66.67	88.54	30.21
N6	52.08	54.17	68.75	28.13
N8	55.21	59.38	80.21	29.17

BEST BOTTOM

Subject	Material	Thickness	Angle	All
N3	51.04	43.75	95.83	20.83
N4	66.67	59.38	67.71	35.42
N6	60.42	64.58	93.75	40.63

BEST AIR

Subject	Material	Thickness	Striker	All
N4	86.46	83.33	59.38	45.83
N7	91.67	84.38	45.83	36.46
N10	85.42	82.29	42.71	35.42

N4 FREE-FIELD

Session	Material	Thickness	Angle	All
3	63.89	58.33	91.67	38.89
4	58.33	61.11	91.67	36.11
6	55.56	55.56	100.00	36.11
Test	53.13	66.67	88.54	30.21

N6 BOTTOM

Subject	Material	Thickness	Angle	All
2	50.00	55.56	75.00	22.22
4	61.11	55.56	83.33	38.89
7	47.22	41.67	88.89	25.00
Test	60.42	64.58	93.75	40.63

N4 AIR

Subject	Material	Thickness	Striker	All
5	88.89	75.00	72.22	52.78
6	80.56	88.89	66.67	52.78
7	88.89	72.22	58.33	41.67
Test	86.46	83.33	59.38	45.83

	Chance	Material	Thickness	Angle/ Striker	All
		50.00	50.00	33.33	8.33
Statistically	Training Sessions	67.00	67.00	52.77	25.00
Significant	Test Session	61.00	61.00	43.75	16.67

Table 7.1-1 Best and Top Performer's Performance and Chance Levels for Sessions in Scaling Runs

	S5M	S5P	S5W	S1M	S1P	S1W	B5M	B5P	B5W	B1M	B1P	B1W
S5M	4	0	0	0	0	0	0	0	0	0	0	0
S5P	16	16	0	0	0	0	0	0	0	0	0	0
S5W	15	12	8	0	0	0	0	0	0	0	0	0
S1M	15	15	16	4	0	0	0	0	0	0	0	0
S1P	16	14	16	16	6	0	0	0	0	0	0	0
S1W	15	11	13	16	15	10	0	0	0	0	0	0
B5M	15	16	15	16	16	16	10	0	0	0	0	0
B5P	16	16	15	16	16	16	15	6	0	0	0	0
B5W	16	15	16	16	16	16	7	14	8	0	0	0
B1M	15	16	16	16	15	16	15	16	16	8	0	0
B1P	16	16	16	16	13	15	16	16	16	15	14	0
B1W	16	15	16	16	15	15	16	16	16	14	12	10

Table 7.2-1 Lower Triangular Dissimilarity Matrix for Air Subject N4

7.3 INDIVIDUAL DIFFERENCES MODEL

The individual differences scaling (IDS) model was chosen to create the multidimensional solutions for the six sets of input data for the top performers described above. The model used a weighted Euclidean distances measure to produce a non-rotatable space in which the placement of the signals was the best fit for all subjects' confusions. The IDS model, unlike other scaling algorithms, produces axes which may not be rotated after the solution is found. This means that the dimensions can be directly interpreted, given the assumption that the scaling model describes the data accurately[3].

7.3.1 Scaling Model

The IDS model took as input matrices of symmetric, dissimilarity data. The Best overall session performers' data were run as matrix conditional, while the single top performers' data were run with an unconditional restriction. The matrix and unconditional indicators simply dictated the way in which responses from matrix to matrix in the input set were treated by the algorithm. Numbers were treated as equal only within matrices for matrix conditional, while the same number was treated equally across matrices for the unconditional condition. For instance, a total of 2 in a matrix for top performer N4 in the Free-field condition was not necessarily the same as a 2 from N6. However, N4's response of 2 in a matrix for the third training session was seen as equal to a

2 in his fourth training session's matrix. The unconditional assumption allowed the scaling algorithm to account for more of the variance in the data. The remainder of the settings for the scaling runs were equal for all subject sets. Solutions were created for two to five dimensions, where one n-dimensional set of data was chosen for analysis.

Although only one set of dimensions was produced per solution, this scaling method allowed for the subjects to use the dimensions differently from one another. In other words, if there were three dimensions provided by the solution, each of the subjects could use the dimensions to a greater or lesser degree than other dimensions or other subjects. The variation of the individual use of the dimensions was represented by a subject weight for each dimension in the solution. Overall measures were also provided by the solution which indicated how the subjects as a group used the individual dimensions.

7.3.2 Subject Weights

For a three-dimensional solution, the subject weights were treated as the coordinates in 3–space of a vector with its origin at (0,0,0). The vectors from each of the subjects could then be viewed relative to one another. It was important to look at the weights as vectors, not as raw weights due to the way in which they are computed by the IDS method. A comparison across subjects of their individual raw weights is not valid, but of the vectors defined by those weights is. For instance, if the weights for two subjects represent points far from one another but along the same vector from the origin, those subjects used the dimensions with the same relative weighting. For the comparison, a method was devised to convert the raw dimension weights to vectors. The vectors could then be compared directly to observe how the subjects used the dimensions differently.

The best method for comparison was derived from knowing the angles from a given subject weight vector to each dimension axis in the solution space. The basis of the angles was the vector produced when the subject weights for dimensions 1, 2, and 3 were treated as the coordinates on the x, y, and z axes. In order to compute an angle, the xyz coordinate from the vector was used in conjunction with each axis individually, and the (0,0,0) point of origin, to form a plane in space. The axis of interest was assigned a point 1 unit from the origin to use as its coordinates. For example, if the angle from the vector to the x axis were desired, the point used on the x axis was (1,0,0). The law of cosines, in Equation 1, was applied to the three points in the plane, and the angle from the subject weight vector to that axis of interest was computed. This was repeated for

each of the other two axes, giving three total angles which then were compared to each other and to other subjects' angles.

(1) $a^2 = b^2 + c^2 - 2bc \cos A$

where: A is the angle between the vector and the current axis; and a, b, and c are the origin, unit point on the axis, and endpoint of the vector, respectively.

The subjects' dimension weights angles directly related how each subject used the three dimensions, where a small angle indicated that the dimension was used substantially and a large angle that it was used less. A vector with equal weights had angles of 54.736° to each of the axes, and thus to the dimensions. A comparison of the subjects weights' angles to the equal weights' angles shows how far the subjects deviated from an "equal" use of the dimensions, and consequently how much the subjects used the dimensions. For instance, the angles for N3 from the Best Bottom solution, shown in Table 7.4.2-1(a), show that dimensions 1 and 2 were used to almost the same degree in classifying the signal parameters, and were close to the equal use, while dimension 3 was used to a much smaller extent. In contrast, N4 used dimension 1 highly, but dimensions 2 and 3 much less. As is shown in these examples, the observation of the angles across subjects and dimensions was a convenient means of discerning the extent to which subjects within one individual differences scaling solution used the dimensions produced.

Another set of measures produced by the scaling solution included a weirdness level for each subject. The weirdness indicated how much the subject's use of the set of dimensions varied from that of the "typical subject." The typical subject's vectors were based simply on the average of the subject weights for all subjects in the solutions. For the weirdness measures to be computed the typical subject's vectors were normalized to orient them along the equal use vectors at 54.7° from the dimensions. The subjects' weight vectors were then normalized in the same manner, and the weirdness index for each subject was computed.

The individual differences model also gave a measure of the relative importance of the dimensions within each solution, which together provided an overall measure of the variance in the original data accounted for by the solution. Given more dimensions, and therefore more parameters, the scaling algorithm could account for more of the variance in the data. In this case, three dimensions were chosen as sufficient to account for the variance in the data for the Free-field, Bottom reflection, and Air data conditions while producing a reasonable number of dimensions for analysis.

7.4 SCALING RESULTS

Two groups of matrices were used as input to the scaling runs for the Free-field, Bottom and Air signal conditions; one consisted of the three test sessions from the Best performers, and one of the overall top performer's three best training sessions plus the test session. The two sets of dimensions produced by the scaling runs for each of the three signal conditions are illustrated and described here. Subject weights which reflect the use of the dimensions in each solution, weirdness measures which show the amount of variance accounted for by each dimension, and the overall importance of the dimensions to the subjects are also detailed here. The dimensions discussed here are related to acoustical measures of the signals and to neural network nodes in Section 10.

7.4.1 Free-Field Condition

The Free-field condition's two sets of scaling dimensions are displayed in Figures 7.4.1–1 and 7.4.1–2. The coding scheme for the signals in the dimension figures here and throughout this section is as follows. The initial letter represents a material of Brass or Steel and the next digit represents a thickness of 10% or 5%. The last character represents either an insonification angle of 90°, 45°, or 0° for the underwater signals or a striker type of Metal, Plastic, or Wood for the Air signals. For example, B10 stands for a target which is brass with a shell thickness of 10%, and is at 0° relative to the transducer. The subject weights for each of the dimensions, shown in Table 7.4.1–1, were an indication of how much the subjects used the dimensions in each session included in the solution.

Five of the six dimensions in the two Free-field solutions break down by Angle to differing degrees. It is particularly interesting to note that Angle is the only parameter that separated readily on any of the dimensions. The first dimensions for the Best three and single best performers separated the 90° signals from the rest. The fact that this occurred on the first dimension where the overall importance level ranged from 0.49 to 0.57 indicates that it was by far the easiest distinction for the subjects to make during the Free-field experiment sessions. The second dimension, which accounts for the next largest amount of variance with importance levels of 0.13 and 0.19, cleanly separates all three Angles, with the single exception of the S54 signal class. The third dimension for N4 also separates the three Angles has an importance rating of 0.17. The exception is the signal class S50 which is widely misplaced at the opposite end of the dimension from other 0°

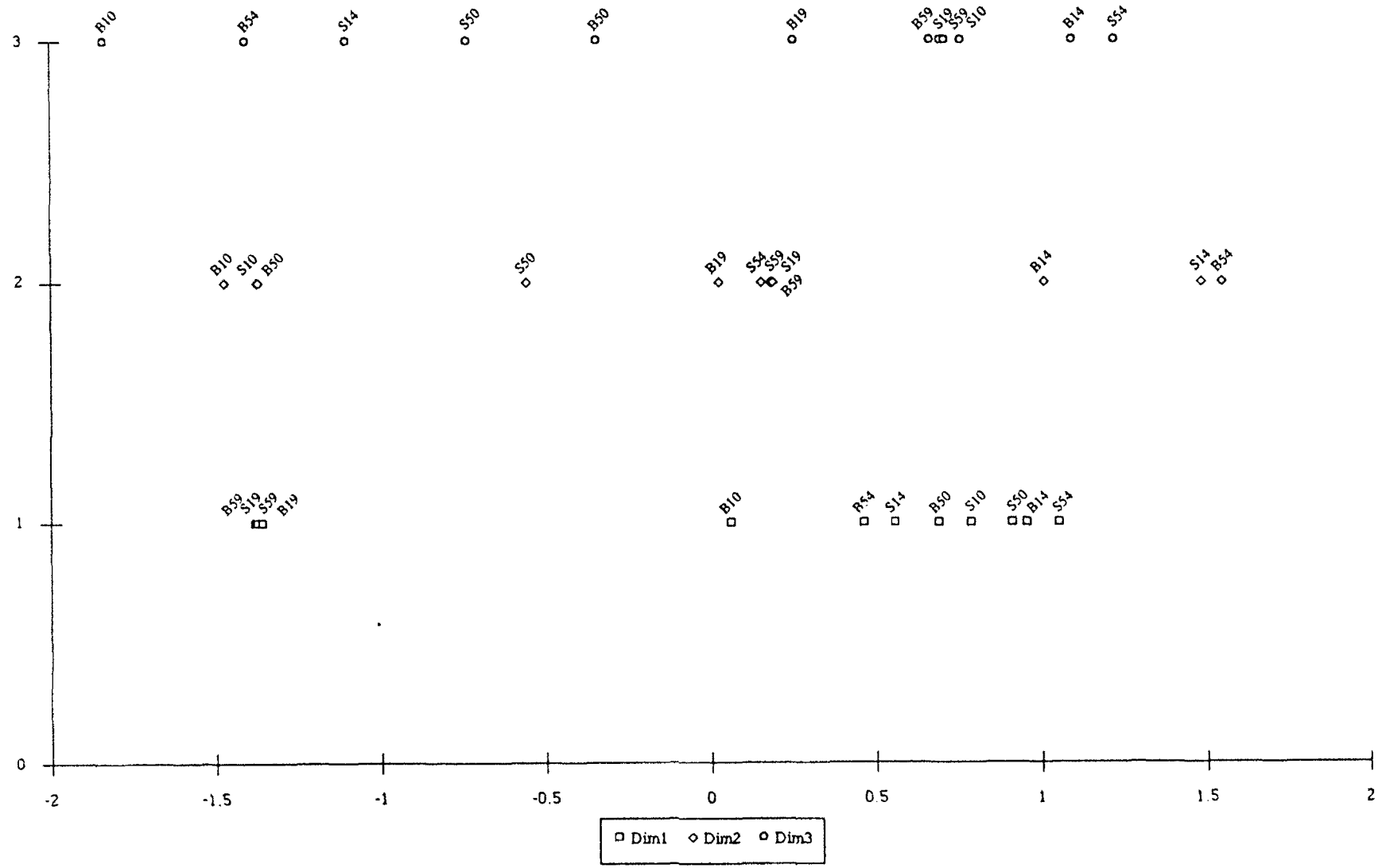


Figure 7.4.1-1 Free-Field Best Subjects' Scaling Dimensions

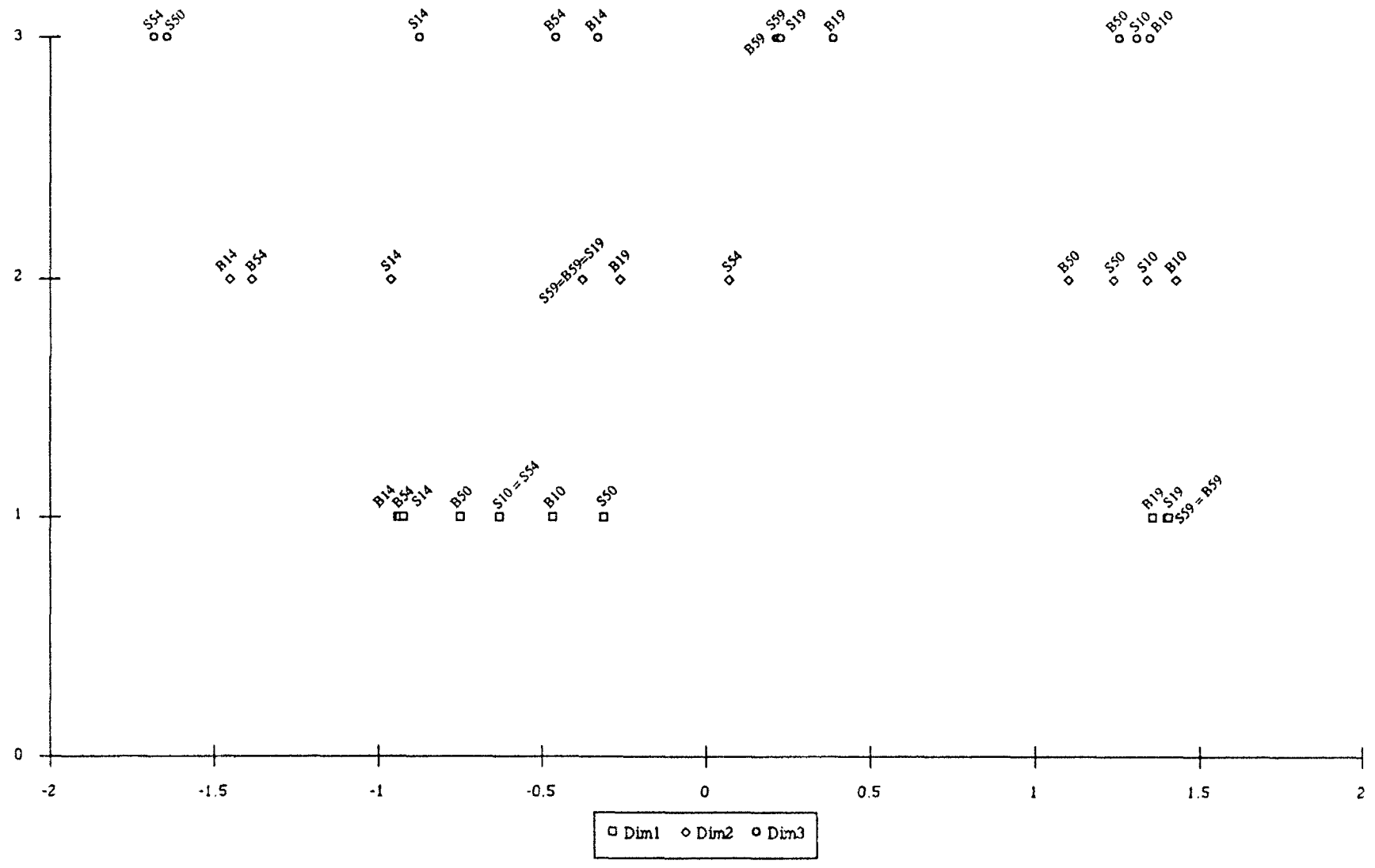


Figure 7.4.1-2 Free-Field Subject N4's Scaling Dimensions

signals. The third dimension for the Best performers, however, did not obviously distinguish any parameter. This dimension also accounted for the lowest level of variance of any of the dimensions. The clustering of the 90° signals in five of the six dimensions points out how similar they sounded to all of the Free-field subjects. Remember that close proximity on a dimensions is an indication of a high degree of confusion. Signals separated by Angle along a dimension, therefore, means that at least some subset of the subjects tended to confuse signals of one angle more with one another than with signals of other angles.

FREE-FIELD

Table 1(a) Best

	Subject Weights			Angles			
Subject	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Weirdness
N4	0.704	0.432	0.173	33.46	59.22	78.15	0.306
N6	0.682	0.118	0.371	29.75	81.33	61.79	0.435
N8	0.704	0.421	0.309	36.59	61.27	69.36	0.113
Overall Importance	Dim1 0.485	Dim2 0.126	Dim3 0.088				

Table 1(b) N4

	Subject Weights			Angles			
Session	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Weirdness
3	0.100	0.057	0.063	40.30	64.37	61.26	0.093
4	0.103	0.071	0.037	38.02	56.83	73.57	0.229
6	0.099	0.080	0.040	42.10	53.13	72.61	0.251
Test	1.500	0.855	0.827	38.41	63.48	64.41	0.026
Overall Importance	Dim1 0.570	Dim2 0.186	Dim3 0.173				

Table 7.4.1-1 Free-Field Scaling Solutions' Usage Measures for Best and Single Top Performers

Several differences that are notable between the scaling runs are revealed mainly in observing the single best performer's solution. N4's judgment of the ninety degree signals is very similar over the three dimensions. The 90° signals are closely clustered in each case and the positions of the classes S59 and B59 are always equal to one another. The 90° signals are closely clustered for the

three dimensions of the Best performers, but not as closely as for N4. This implies that N4 had a higher level of confusion among the 90° signals than did the Best subjects. Conversely, N4's clustering of the 0° and 45° signals on the first dimension shows a better separation of these angles than the Best performers. The 45° signals, except for the troublesome S54, separate from the 0°, and both sets of angles are distinguished highly from the 90° signals. The third dimensions, as mentioned, are also quite different from one another.

The subject weights produced by the scaling model are interesting to examine for possible relationships with performance levels. The outstanding characteristic of the weights for the Best solution is that N6 uses dimensions one and three to a noticeably greater extent and dimension two much less than subjects N4 and N8. The difference in his approach is also reflected in the relative magnitude of his weirdness (0.43 vs. 0.31 or 0.11). The second dimension is where the Angles were broken out completely, and lack of use of this dimension is reflected in N6's relatively low performance on Angle, as seen in Table 7.1-1. Although his performance is below that of his peers, it is still above the statistically significant level of 61% correct. N4 is slightly better than N8 for Angle, and there is a probable relationship to the fact that N4 uses dimensions one and two, which separate by Angle, more than N8 does. Although there is a noticeable difference in N4 and N8's Thickness performance, it cannot be directly related to use of any of the dimensions since none of them broke down by that parameter. It is interesting that N4's performance levels were higher than N8's although his weirdness was also higher. This dismisses the tendency to assume that a higher weirdness, and thus distance from the typical subject's use of the dimension, implies that the performance will be lower for the parameter which the dimension represents.

The subject weights for N4 in the single best performer's solution do not reveal as many possible correlations as those found in the Best solution's weights. Here, and in the single best solution for Bottom and Air, each of the matrices represents a particular session of the experiment so it will be referred to as such. This is opposed to a matrix from the Best solution being referred to by the subject whose data it contains. The matrices for sessions 4 and 6 show that the dimensions were used in a very similar fashion in the two sessions. The same holds true for the dimension use in session 3 and the test session, although the use by the two pairs of sessions is not the same. The weirdness measures for the four sessions parallel the dimension use levels. Unfortunately, the performance levels for the parameters show no direct association with the dimension use. On the other hand, the high Angle performance of 89-100% is reflected in the fact that all of the dimensions directly deal with the Angle differentiation. The presence of such an effect on all

dimensions implies that the Angle parameter was consistently emphasized throughout N4's performance.

7.4.2 Bottom Reflection Condition

As in the case of the Free-field condition, five of the six dimensions in the two Bottom scaling runs separated according to the signals' Angles to differing extents. The dimensions are shown in Figures 7.4.2–1 and 7.4.2–2. Dimension 1 has the same 90° vs. 45°/0° separation for both scaling solutions as in the Free-field condition. The second dimension in both cases also matches the Free-field solutions in distinguishing each Angle separately. In fact, the order along the second dimensions for the two Bottom solutions is nearly identical. There is also a slight separation for Thickness within Angle groupings, particularly for the single top performer, N6. Dimension three in N6's solution also is separated by Angle, but in a different manner than usual which is discussed below. As for the Free-field's Best performers, the third dimension for the Bottom condition's Best performers does not readily distinguish any of the three parameters.

The prevalence of the Angle parameter in five of the dimensions is reflected in the high performance for Angle across the subjects. Again, as for Free-field, the variance in the data accounted for by the first two dimensions in both Bottom solutions, which break down by Angle, is very high. The first dimensions account for the most at levels of 0.48 and 0.598, while the second dimensions have significant levels of 0.23 and 0.27. Neither third dimension has a very high level of importance at 0.08 and 0.07.

N3 of the Best performers did very well with the Angle parameter, and relatively well overall. The remaining parameters of Material and Thickness, however, he did not distinguish well. N6 identified Material and Thickness significantly better, and his overall performance was almost double N3's. Oddly enough, their subject weights, and thus their dimension use, was very similar. The weights, shown in Table 7.4.2–1(a). Dimensions one and two, viewed from a 3–dimensional perspective, show some Thickness separation within the Angle categories. The assumption is that since the Thickness separation is not as obvious, N3 did not pick up on the subtlety of the Thickness differentiation, but concentrated on Angle separation. N6, in addition to his high performance on Angle, used the same dimensions similarly, but was able to discern more subtle features of the signals, and was able to achieve superior performance. Despite the differences in their performance levels, N3 and N6 had similar weirdnesses.

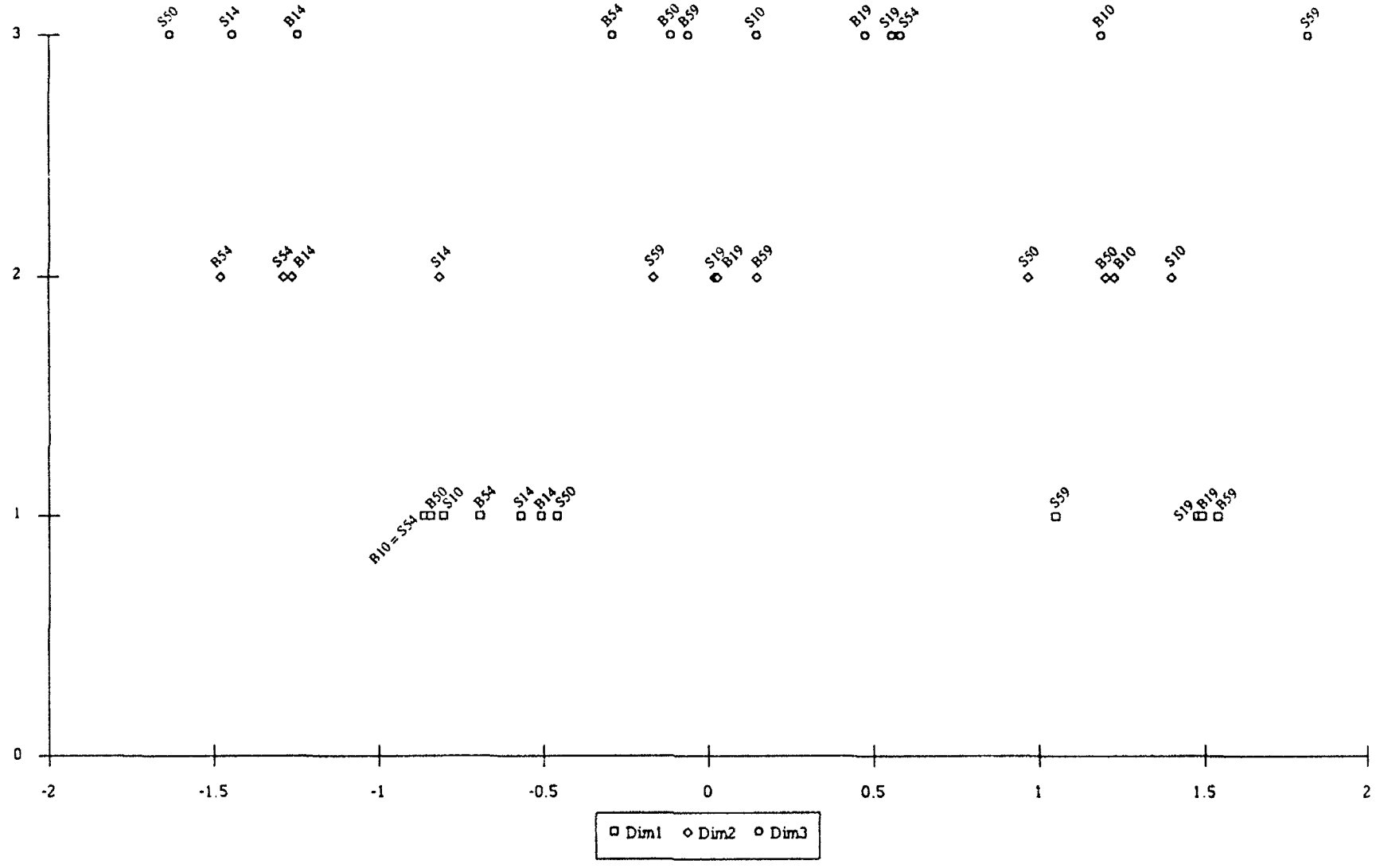


Figure 7.4.2-1 Bottom Best Subjects' Scaling Dimensions

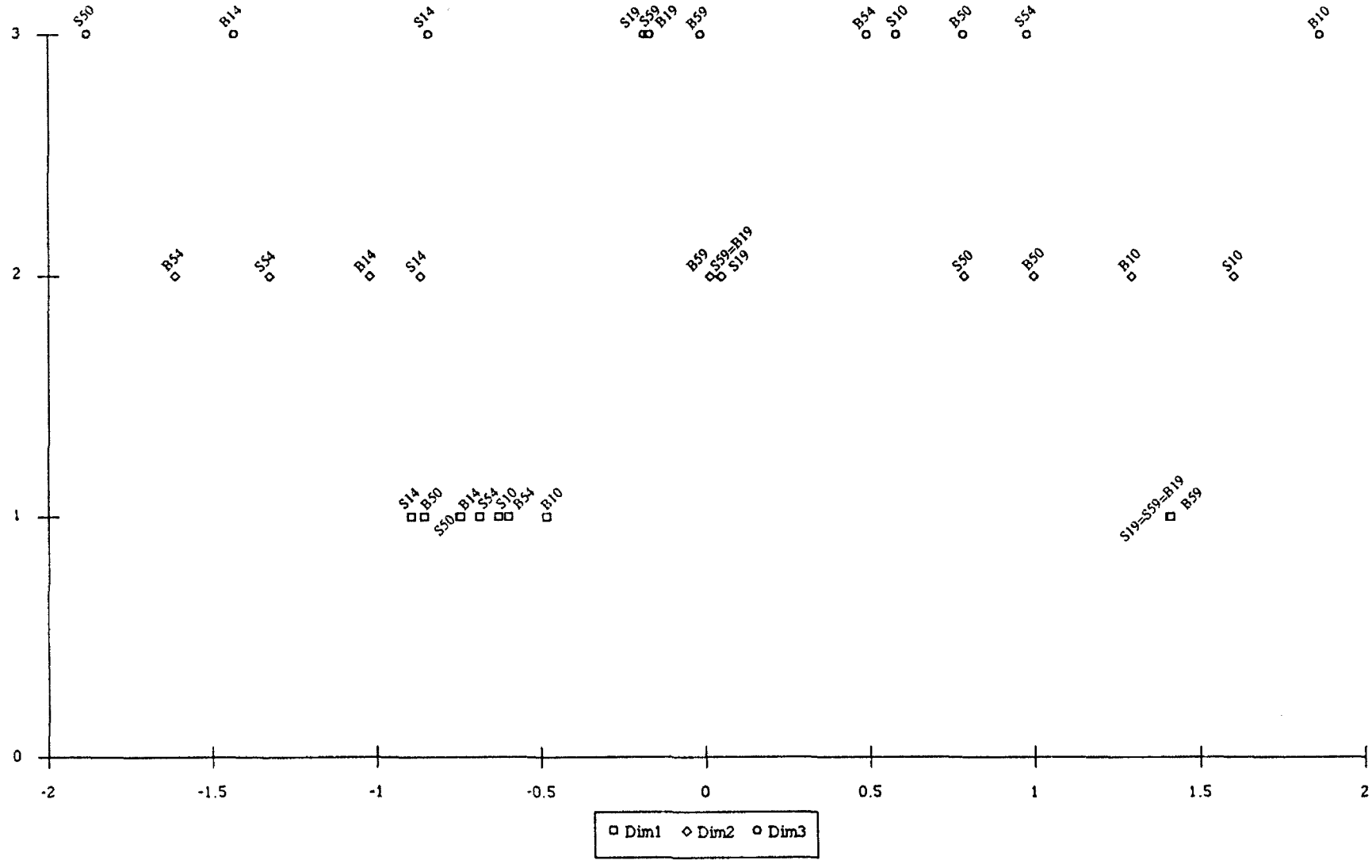


Figure 7.4.2-2 Bottom Subject N6's Scaling Dimensions

BOTTOM

Table 1(a) Best

	Subject Weights			Angles			
Subject	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Weirdness
N3	0.665	0.608	0.286	45.26	49.99	72.42	0.192
N4	0.739	0.114	0.260	20.96	81.70	70.89	0.452
N6	0.676	0.552	0.300	42.89	53.28	71.02	0.128

Overall	Dim1	Dim2	Dim3
Importance	0.482	0.229	0.080

Table 1(b) N6

	Subject Weights			Angles			
Session	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Weirdness
2	0.075	0.040	0.019	30.61	62.49	77.63	0.129
4	0.074	0.041	0.022	32.10	62.12	75.39	0.076
7	0.071	0.047	0.024	36.72	57.78	74.30	0.006
Test	1.542	1.031	0.521	36.84	57.65	74.30	0.008

Overall	Dim1	Dim2	Dim3
Importance	0.598	0.267	0.068

Table 7.4.2-1 Bottom Scaling Solutions' Usage Measures for Best and Single Top Performers

The Best solution's subject N4 stood out in his greater use of dimension 1 and greatly decreased use of dimension 2 compared to N3 and N6. This corresponds to his inability to separate the 0° and 45° from one another, although he could easily distinguish both from the 90° signals. His weirdness level at 0.45 was also much higher than that of the other subjects. Although his Angle success level was only 68%, as compared to 94-96% of subjects N3 and N6, it was still significantly above chance levels. This is due to his excellent identification of the 90° signals, and chance performance on the 45° and 0° signals. Ironically, considering his relative performance on Angle, N4's performance on Material was the highest of the three Best performers at 67%. This also is reflected in his unique use of the dimensions, particularly his lack of stress on the second dimension where Angle is the most important parameter. It is apparent from N4's performance that Material is distinguishable to some extent, although there is no obvious breakdown for Material on any of the dimensions.

The scaling solution using only N6's data shows again that Angle is the most easily determined parameter. The first two dimensions have a clear Angle separation, while the third dimension separates the 90° signals from the others, but in a more unusual manner than has been seen until now. The signals on either side of the 90° signals do not all fall into either the 0° or 45° category. Observation of dimensions 2 and 3 together in a 3–space perspective, however, shows that the Angles separate well with S50 as a slight problem.

As the sessions progress there is an overall decrease in the use of dimension 1, an increase in the use of dimension 2, and a slight increase for dimension 3, and this is shown in Table 7.4.2-1(b). This change shows the parallel between dimension 2's complete separation on Angle, the 90° signals' placement in the middle of dimension 3, and the rise in N6's performance for determining the signals' Angles. An increase in his use of dimension 2, with its perfect separation of Angles, shows that N6 is more able to make the fine discriminations shown by the dimension. Additionally, the 0° and 45° signals are separated by Thickness on dimension 2. The dimensions show that the 10% signals within each Angle are separate from the 5% signals. This separation is reflected in the expected increase in performance for Thickness as the use of dimension 2 increases. The increase in performance occurs, with an exception to the trend at session 7 which can be explained by observing what happens to the Angle performance. In session 7 Angle is the only parameter on which performance improves over the levels from previous sessions, while the other levels fall a noticeable amount. The theory is that the subject concentrated on improving his Angle discrimination ability at the expense of the other parameters. The test session performance levels show, however, that he is competent for Material, Thickness and overall identification of the signals, and has returned to the previously increasing trends in performance and dimension use evident in sessions 2 and 4.

Although session 7's performance is a marked exception to that of the other sessions, the weirdness for it is very small at 0.006. Other sessions' weirdnesses only range up to 0.13, which itself is small, but it would be expected that the weirdness would be highest where the performance trends varied the greatest amount. This is not the case, however, and it may be attributed to the fact that none of the weirdness levels was particularly high for any of the four sessions in the solution.

7.4.3 Air Condition

The task of distinguishing the parameters for the Air signals is fundamentally different than doing so for the Free-field and Bottom cases. This difference is readily reflected in the scaling solutions for the two sets of matrices from Air subjects. Where the Free-field and Bottom solutions showed many divisions for Angle, but few for Material and Thickness, the Air scaling solutions are separated mainly by Material and Thickness, with some distinctions for Striker. The first dimensions for both scaling runs divide by Thickness, the top performer's perfectly, and the Best's with two exceptions. Similarly the third dimensions separate perfectly by Material with only one exception in the Best solution. The second dimensions are not perfect, but each has partial separations for Material and Thickness, and the top performer's second dimension separates by Striker to some extent as well. The dimensions for the Air solutions are displayed in Figures 7.4.3–1 and 7.4.3–2.

The scaling results show an affinity of the Best subjects for determining the Material and Thickness of the Air signals. The 82-92% success rate for these parameters by all three subjects is well above the statistically significant level of 61%. The signal distribution along the dimensions parallels the performance on the two parameters. The first dimension has the 5% and 10% signals widely separated, with the exception of the S1M and S5W classes. The second dimension has a diverse clustering of signals, with some cases based on Material, and others on Thickness. The Brass 5% signals are at the extreme lower end, five of the six 10% signals cluster in the middle, and four of the Steel signals are toward the high end of this dimension. The third dimension separates cleanly by Material, with the exception of the S5M class. There is also a Thickness differentiation among the Brass signals, with the 10% signals at one extreme and the 5% signals toward the middle of the dimension where the Steel 5% signal class is also included.

A plot of Best dimensions two vs. three, seen in Figure 7.4.3-3, shows that a perfect Material separation exists about the boundary between the positive and negative quadrants. A good separation for Thickness is also incorporated into the Material distinctions in this view, with only the two exceptions which were apparent in the first dimension. In other words, the S1M, S5W confusion seen on dimension one is also present in the 2–dimensional view of the dimensions two and three. Overall, the dimensions separate very cleanly for both Material and Thickness, and this is reflected in the performance levels. Ironically, there is little indication of visual separation for the Striker parameter on any of the three dimensions. Regardless of this, the subjects'

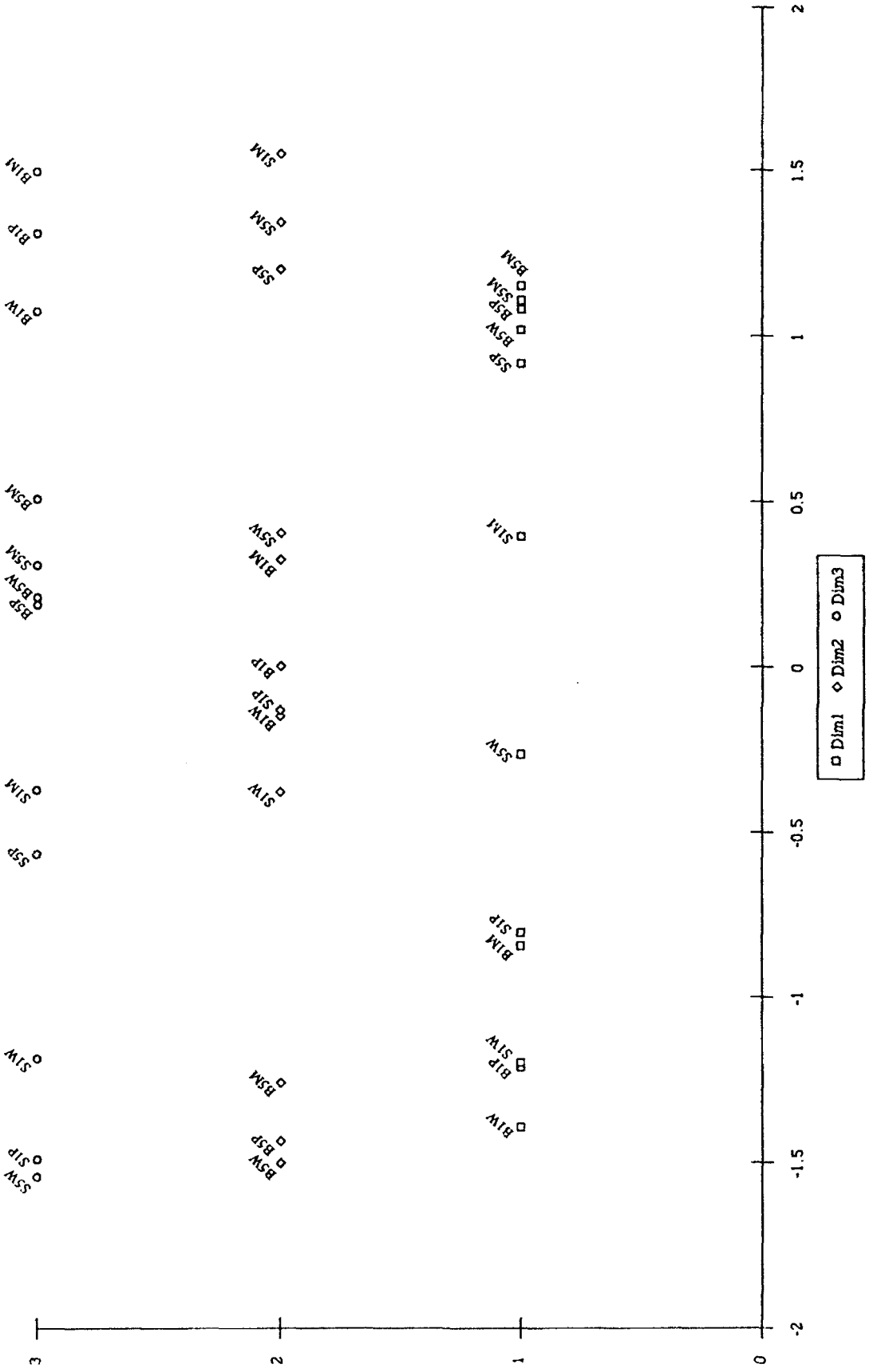


Figure 7.4.3-1 Air Best Subjects' Scaling Dimensions

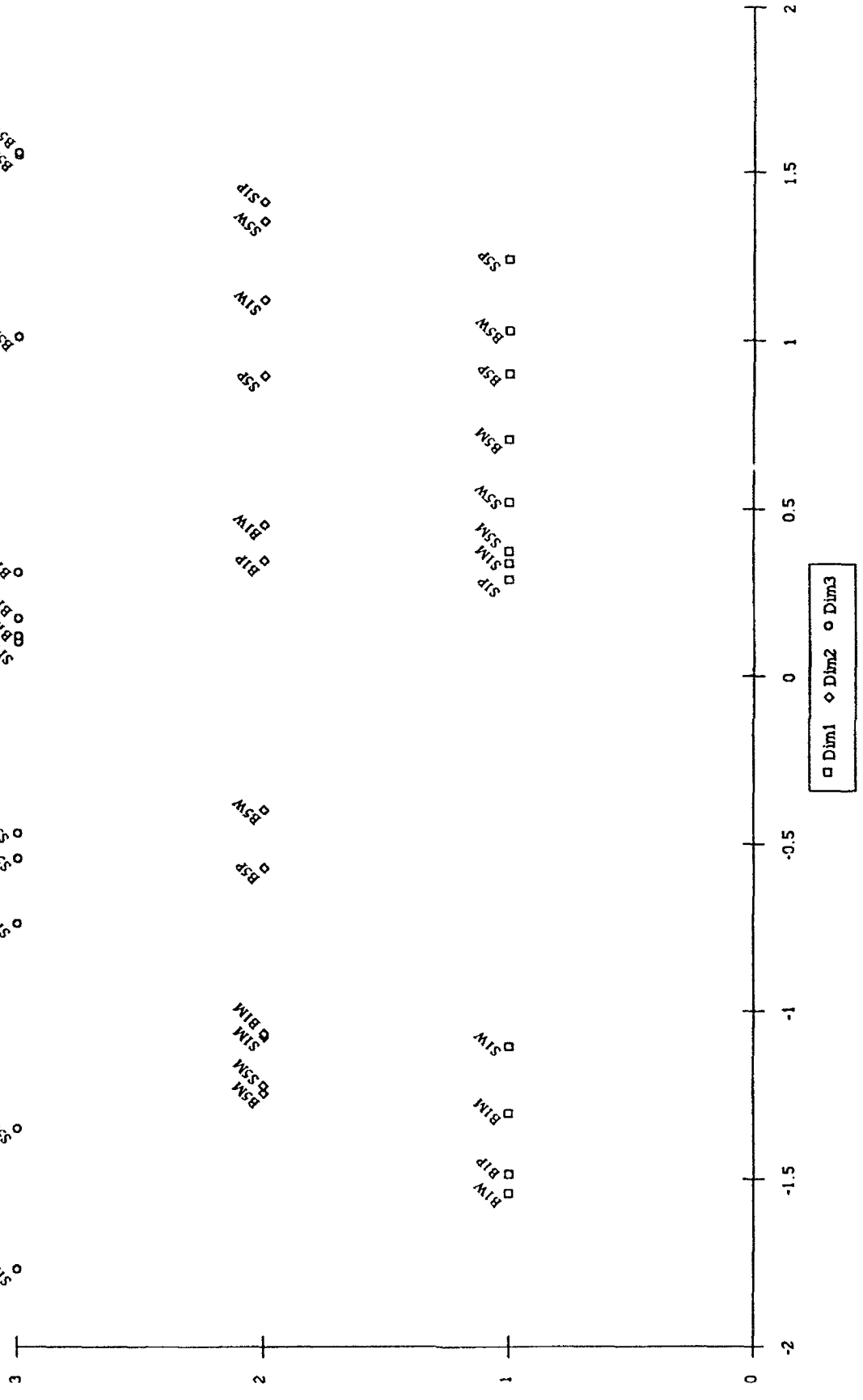


Figure 7.4.3-2 Air Subject N4's Scaling Dimensions

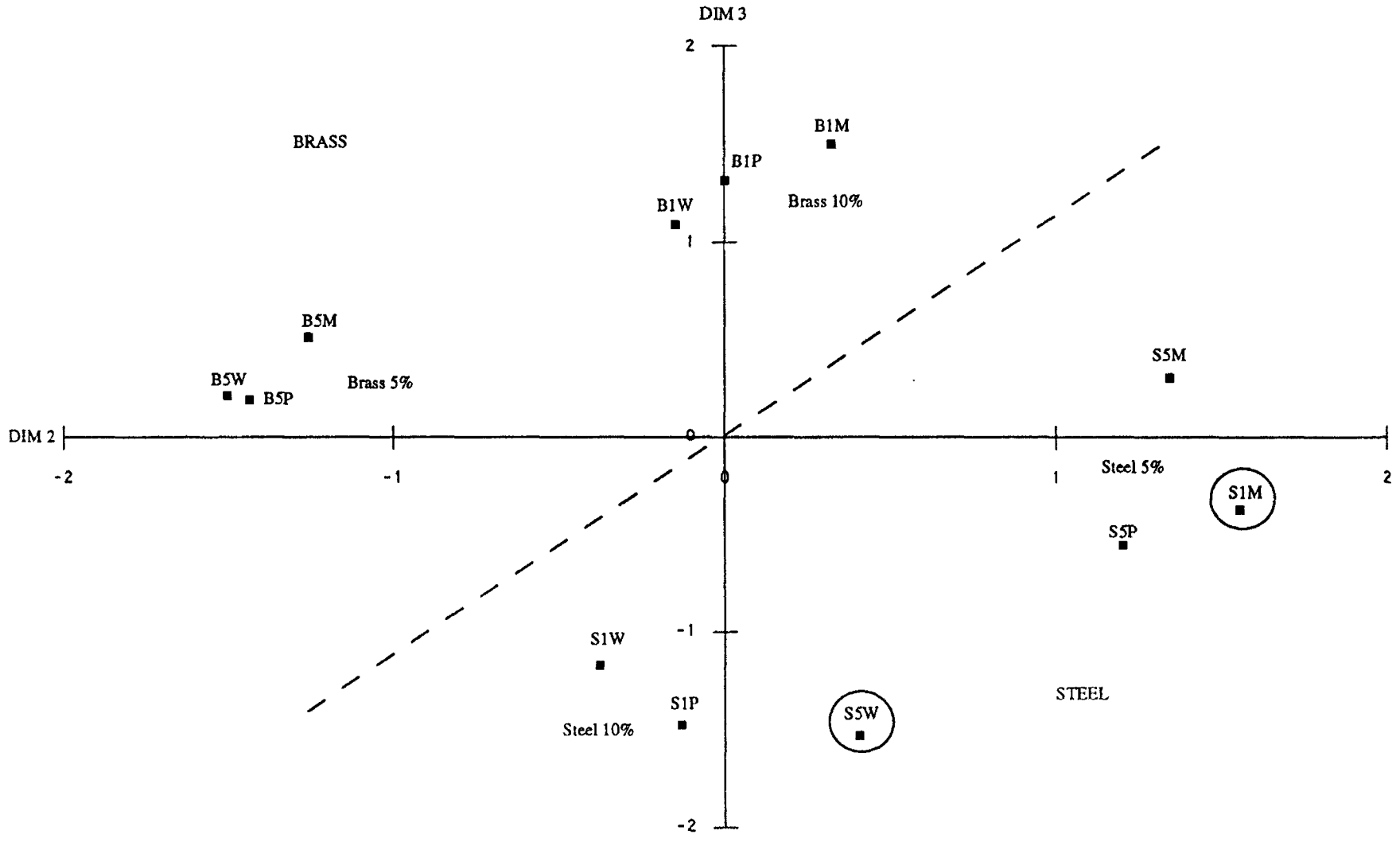


Figure 7.4.3-3 Air Best Subjects' Dimensions Two vs. Three

performance levels of 43-59% for Striker are at or near the statistically significant level of 43.75%. Their overall performance levels of 35-46% are also well above 16.67% which is where the performance is judged to be statistically above chance.

The subject weights produced by the individual differences model, seen in Table 7.4.3-1(a), show no significant correlation between individual subject's performance and their use of the dimensions. The dimensions were used by the subjects almost equally both relative to one another and across other subjects. The weirdness levels for the subjects also reflect this consistency, and only range from 0.03 to 0.05. These results made it difficult to associate any particular performance behavior with using a given dimension or set of dimensions.

AIR

Table 1(a) Best

	Subject Weights			Angles			
Subject	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Weirdness
N4	0.463	0.404	0.351	49.14	55.19	60.26	0.045
N7	0.539	0.534	0.455	52.47	52.87	59.05	0.030
N10	0.432	0.399	0.402	52.67	55.94	55.64	0.053

Overall	Dim1	Dim2	Dim3
Importance	0.231	0.203	0.164

Table 1(b) N4

	Subject Weights			Angles			
Session	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Weirdness
5	0.062	0.061	0.054	52.95	53.37	57.98	0.073
6	0.065	0.066	0.049	51.61	51.05	62.09	0.116
7	0.063	0.059	0.056	52.32	55.16	56.78	0.054
Test	1.260	1.024	1.002	48.67	57.54	58.31	0.012

Overall	Dim1	Dim2	Dim3
Importance	0.400	0.265	0.253

Table 7.4.3-1 Air Scaling Solutions' Usage Measures for Best and Single Top Performers

The consistent occurrence of Material and Thickness separation along all three dimensions, the performance on these parameters, and the equal use of the dimensions are paralleled by the overall

importance placed on the dimensions. Unlike in the Free-field and Bottom conditions, the variance importance levels here range only from 0.16 to 0.23. The relatively small difference among dimensions emphasizes that all of the dimensions were used by the subjects in their classification decisions, particularly for Material and Thickness.

The first dimension for the single top performer for Air, N4, separates perfectly by Thickness, although S1M and S1P are separated from the other 10% signals and are near the 5% signals. This does not mean, necessarily, that N4 could not distinguish the S1M and S1P classes of signals, only that they were confused with the 5% signals more often than with the other 10% signals. Dimension two has an interesting array of signal clusters. The Plastic and Wood strikers consume three quarters of the dimension with the Metal strikers clustered in the lower quarter. The separation of the Metal signals was an important result and it was reflected in N4's superior Striker performance of 58-72% over the other two Best subjects' levels of 43-46%. Within the Plastic and Wood distribution the Steel, Brass 10%, and Brass 5% signals are grouped separately. Within the Metal cluster the 10% and 5% signals are separate. The different groupings on this dimension encompass all three parameters to varying extents. The third dimension is equally mixed across the three parameters. Overall it is separated perfectly by Material. Within the Brass signals the 5% and 10% signals are separate, and within the Steel signals the Metal signals are grouped separately from the Plastic and Wood signals.

Overall the dimensions divide well by Thickness and Material, but only separate Striker as Metal vs. Plastic/Wood. This difference is reflected in the performance for the three parameters. N4 has a success rate of 81-89% for Material, 72-89% for Thickness, but only 58-72% for Striker. Despite this, his Overall performance is well above chance levels of 16.7% and 25% for the training and test sessions respectively.

N4's use of all dimensions is shown in the small difference in the amount of variance accounted for across dimensions. The levels ranged only from 0.25 on the third dimension to 0.4 on the first. The closeness in the range stresses that the information represented on all dimensions contributed significantly to N4's performance of the classification task.

As was the case for the Best performers the subject weights for N4's solution, shown in Table 7.4.3-1(b), are relatively consistent. This consistency implies that the dimensions were weighted, and thus used, approximately equally across sessions. In N4's case, however, his performance

on Thickness is reflected in his use of dimension one which separated Thickness perfectly. Specifically, as he uses dimension one more, his Thickness performance level increases. Similarly, his performance for Material parallels his usage of dimension three which was predominantly separated by Material. Although dimensions two and three break down somewhat by Striker the performance trends for Striker are not exhibited in those dimensions' subject weights. The consistency across the individual sessions' use of the dimensions is also shown in their weirdness values which are of small magnitude and range from 0.01 to 0.11.

7.5 SUMMARY

Overall the scaling solutions provided dimensions, and other weight-related measures, which were used in later analyses to derive signal features used by the humans in performing the classification tasks for the Free-field, Bottom, and Air signal conditions. The Free-field and Bottom solutions exhibited the subjects' predominant ability to separate the signals by Angle. These subjects were especially accomplished at separating the 90° signals from the group of 45° and 0° signals. The Air solutions contained more diversity for all three parameters, but showed that the subjects were particularly adept at discerning Material and Thickness. Many of these performance results were reflected in the subjects' use of the dimensions, which was shown by examining the subject weights for each of the dimensions alone and together. The discussion of the scaling solution dimensions, the signal classes' distribution over them, and the subject weights associated with them is only a portion of the evaluation of how the humans went about discriminating signal parameters. The signal features which presumably formed the basis of the subjects' processing are explored when correlations between the human and network data, as well as signal parameterizations, are examined in Sections 9 and 10.

8.0 NEURAL NETWORK TRAINING

The neural network experiments determined the ability of networks to classify sonar returns in the frequency and time domains, as well as frequency over time, under a certain set of training parameters. These experiments provided network hidden nodes and data to use in multidimensional scaling routines, the results of which could be used for comparing the processing strategies of networks and human subjects performing the same signal classification task. Of the many possible neural network architectures, both the backpropagation and the integrator gateway networks were chosen as the models to use. Initial studies with the counterpropagation network architecture and training regime indicated that the method was not suited to producing networks with comparable strategies to those of human subjects.

8.1 BACKPROPAGATION

The training schedule for the backpropagation network (BPN) model including training networks with several forms of input data. The trained networks were then tested against signals under differing conditions. The signals used as input were in either the time or frequency domain. The first set of training used input signals in their original "clean" format. In other words, no type of noise was added to the signals as they were fed into the network. These "clean-trained" networks were tested against the original clean signals and signals which had pseudo-random noise added to them. After the clean networks were trained and tested, BPNs were trained with the signals which had pseudo-random noise added to them. For simplicity these signals are referred to as noisy signals in this section and the remainder of the report. The noise-trained networks were then tested against both the clean and noisy signal sets. The results from these networks are discussed and compared later in this section.

In an effort to use concise references to specific networks and nodes within them, the following conventions subsequently will be used. The first portion of the abbreviation refers to the signal set (Air, Free, Bot). This is followed by the number of hidden nodes and which random seed was used. For example, 6H(3) means six hidden nodes, with the third random seed used. Next follows the letter "F", for frequency domain, or the letter "T", for time domain. The domain indicator is followed by the letter "N" if the network was trained with noisy signals; if trained with clean signals, no letter is included. With this notation, all of the network parameters are clearly specified. For example, the abbreviation "Air2H(2)FN" denotes an air signal, two hidden node

network, in the frequency domain, trained with noise from the second random seed. A free signal, zero hidden node, time domain network trained without noise and with the first random number seed would be abbreviated "Free0H(1)T".

To specify nodes within a network, a dash followed by one to four characters is used. For input and hidden layers, this is the character "I" or "H", followed by a number indicating the node. The output nodes are denoted by the following scheme: "B" for brass, "S" for steel, "Ten" and "Five" for ten percent and five percent target thickness, and "0", "45", or "90" for the target orientation in degrees. For example, "Bot4H(1)TN-Ten" refers to the ten percent output node of the Bottom four hidden node network, trained with time domain noisy data using random seed (1). "Air4H(2)F-I7" is the seventh input node for its specified network. Such abbreviations are used for the remainder of the report.

8.1.1 Signal Preparation

To implement the backpropagation networks effectively, it was desired to make the input layers as small as possible, while still adequately representing the information in the signals. This required greatly compressing the signals from their original sizes of hundreds or even thousands of time series points. The practical upper limit on input layer size, in both the time and frequency domains, was approximately fifty. The exact sizes chosen varied, depending on details particular to the signal set and domain. The process of rendering initially very long signals as network inputs will be discussed in two stages: preprocessing and compression. The steps in each stage are described below and summarized in Figure 8.1.1-1.

8.1.1.1 Preprocessing

The preprocessing performed on the Free and Bottom mean 0 adjusted signals, described in Section 4, paralleled the preparations of these signals for the human subjects. The same preprocessing was performed for both the time domain and frequency domain signal compression. A Fast Fourier Transform (FFT) was applied to the mean 0 adjusted 2048 point signals, they were band-pass filtered, and inverse transformed. The ranges of the filter were the same as those used in preparing the signals for the human subjects. The Free-field signals were aligned by the onset of the specular and both the Free and Bottom signals were normalized to the range (0.0, 1.0). The Air signals were subjected to no processing prior to the signal compression.

Preprocessing

Signal Set	Free-Field	Bottom	Air
Original Signal Size	2048	2048	≤ 32456
Band-Pass Filtered	243.2 to 587.9 kHz	229.5 to 587.9 kHz	Not Performed
Means of Alignment	Onset of Specular	Back of Box Return	Impact of Striker

Time Domain Compression

Signal Set	Free-Field	Bottom	Air
Padded/ Truncated Size	800	1333	32768
Window for Averaging	32 Time Points	31 Time Points	1024 Time Points
Final Size of Input	25	43	32

Frequency Domain Compression

Signal Set	Free-Field	Bottom	Air
Padded/ Truncated Size	2048	2048	32768
Hamming Windowed and FFTed	Yes	Yes	Yes
# of Independent Bins	1025	1025	16385
Bandwidth Per Bin	0.9766 kHz	0.9766 kHz	0.9766 Hz
1st Net Input Contains	Bins 249 - 264	Bins 235 - 249	Bins 0 - 512
# Bins in Other Inputs	16	16	512
Final Size of Input	22	23	32

Figure 8.1.1-1 Signal Processing Summary for Network Inputs

8.1.1.2 Compression for Time Domain Signals

The steps involved in compressing the signals varied, depending on the domain and the signal set. In every case it was necessary, at some point, to reduce the size of the signals, and this was always accomplished in the same way. The absolute values of the first N points in a signal were summed and divided by N to make the first input, the next N were used in the same way to create the second input, etc. This process will be referred to below as "averaging the signal over a window of size N." The resulting representation consisted of a factor of N fewer points, but contained information from all the original signal values. Because the absolute values were used instead of a signal's signed values, the result was a good representation of the signal's shape.

8.1.1.2.1 Free-Field Signal Set

For the human subjects, the Free-field signals were cut off to different lengths to reduce any spurious cues present in the noise following the end of the target energy. In this particular context "noise" is used to refer to the energy present in the signal which is not attributable to energy reflected from the target. It would be desirable to do reduce the signals to different lengths for network inputs as well, but because each signal had to be applied to the same input layer, all the signals had to be cut to the same size. Prior to the frequency domain interpolation, the longest Free-field signal prepared for the human subjects was 800 points (the 100 point ramp was started at input 700), so this was the initial length for all of the network signals. No ramp was applied to the signals to be used for network inputs. It was decided that the window size used for averaging in this case would be N = 32, which resulted in an input layer size of 25. This was chosen because it was less than the upper limit of 50, but still contained all the essential features of the signals' envelopes. The final step was to normalize the inputs to the range (0.0, 1.0) again to assure a consistent level for the signals across the input set.

8.1.1.2.2 Bottom Signal Set

Precise alignment of the Bottom signals was not important for the human subjects, due to the periods of silence separating successive sounds during the experiment sessions. However, the nature of network inputs required that some alignment be performed. If the inputs were not consistently aligned within a class, the networks would either fail to learn to classify the signals,

or, more likely, they would learn to distinguish the signals based on individual signal's anomalies. If signals from different classes were aligned improperly, the alignment itself might provide a spurious cue, leading to a non-generalized network solution.

It was a simple matter to align the Free-field signals, due to the consistent and obvious onset of their speculars. The Bottom signals, however, characteristically built up gradually over time, with no obvious or consistent starting point. Fortunately, these signals did possess a well defined "stopping point." In addition to the return from the sandy bottom and the target, each Bottom signal contained a reflection from the back edge of the box in which the target was placed. Although small, this reflection was easily identified in each signal, because it occurred after most of the actual bottom return had decayed. Since the distances between the back of the box and the transducers were constant for all targets and orientations, the reflection from the back of the box provided a stable and consistent marker for the end of each Bottom signal. It was found that, within each signal class, the position of the return from the back of the box was constant across all instances. The position of the back of the box return was therefore determined for each class from the averaged signal.

Once established, the position of the reflection from the back of the box was used as the cutoff for the Bottom signals. It was then empirically determined that even the longest Bottom return was comfortably contained within approximately 1350 points prior to this cutoff. The signal length was then set to 1333 points, which yielded 43 signal inputs after averaging over a window of size 31.

8.1.1.2.3 Air Signal Set

The Air signals were the most straightforward to process since they required no filtering. As was true of the Free-field signals, Air signals were of different lengths for the human subjects. To render them in a form palatable to the networks, they were all made to be the same length. The longest human experiment signal was 32456 points. For processing convenience, this was rounded up to 32768 points for the network inputs. This resulted in no significant change to the information contained in the signal, due to the extremely small values of the signal in the end region. The value 32768 was chosen so that averaging over a window of N = 1024 would produce network inputs of 32 points.

8.1.1.3 Compression for Frequency Domain Signals

The network inputs in the time domain each represented the averaged amplitude of the signal over the period of time spanned by each input point. By analogy, in the frequency domain, it was necessary to create network inputs which represented the averaged amplitude of the frequency components in the band spanned by each input bin. Many of the steps to obtain this goal were the same for the three signal sets. The first step was to take an FFT of each real-valued signal. To facilitate this, the time domain signals just described were zero-padded to make the Free-field and Bottom signals 2048 points long, and the Air signals 32768 points long. The signals then had a Hamming window followed by an FFT applied to them. The results in each case were complex-valued frequency domain representations with as many bins as there were points in the zero-padded time domain signals. These frequency domain representations were converted into complex polar form, yielding an amplitude and phase for each frequency bin. Due to symmetry, many values in the FFT of a real-valued signal are redundant. If the FFT consists of N bins of frequency amplitude data, the amplitudes in bins N/2+1 through N-1 are the mirror image of the values in bins 1 through N/2-1. This means that the FFT may be completely represented by the first N/2+1 independent bins which include the DC offset of the signal, N/2-1 frequency values and the Nyquist frequency value. Only the amplitudes were needed to create the network inputs, so the phases were subsequently ignored.

The acts of performing the FFT and using only the amplitude from each bin thus reduced the size of the frequency domain representations of the signals by almost a factor of 2. At the conclusion of these first steps, the Free-field and Bottom signals consisted of 1025, and the Air signals of 16385, positive values. Following this, the only remaining step was averaging the signals over the appropriate window sizes. The details of how this was performed differed by signal set, and will be described separately below.

8.1.1.3.1 Free-Field Signal Set

Because the Free-field signals were previously band-pass filtered, their FFTs consisted of all zeros outside of the bins containing the frequencies passed. The passed bins were 249 through 601, inclusive, which represent the frequency range 243.2 through 587.9 kHz. (The Nyquist frequency is 1000 kHz; bins 1 through 1025 divide up the range 0 to 1000 kHz, giving 0.9766 kHz per bin.) This range consists of 353 bins; averaging with a window of n = 16 would give 22 inputs, with

one bin left over. The extra bin was simply included in the lowest frequency average, so that the first input actually represented 17 bins (16.602 kHz), starting at 243.2 kHz. Each subsequent input then represented 16 bins (15.625 kHz). The odd bin was included in the first input because the upper limit of the band-pass filter is the same for Bottom signals. By including the extra bin in the first average, the rest of the inputs cover the same frequency ranges as most of the Bottom signal inputs. Table 8.1.1.3.1-1 gives the final correspondence between bins and frequency ranges for the Free-field signals.

8.1.1.3.2 Bottom Signal Set

In all respects, the compression of the Bottom signals was accomplished in the same way as for the Free-field signals. The only difference in the way the two cases were handled was that the lower limit of the band-pass filter in the Bottom signal set was 229.5 kHz, corresponding to bin 235. The total number of bins to be compressed was then 601 - 235 = 367. Averaging over a window of N = 16 would give 22 inputs, with 15 bins left over. Rather than include 15 extra bins in the first average, these bins were averaged to provide one extra input, giving the Bottom compressed signals a total of 23 inputs. The first of these averaged bins represented 15 bins (14.648 kHz), starting at 229.5 kHz, and the rest each represented 16 bins (15.625 kHz). The last 21 of these represent the same frequency ranges as the last 21 of the Free-field signal inputs. Table 8.1.1.3.2-1 gives the final correspondence between bins and frequency ranges for the Bottom signals.

NOTE: The DC offset was not included in creating the frequency domain Free-field and Bottom signals since it had already been set to 0 in the first step of processing the original signals.

8.1.1.3.3 Air Signal Set

The signal-to-noise ratio of the Air signals was so high that they were not band-pass filtered at all. They were simply averaged over a window of 512 bins, with bin 0 (the DC offset) being included in the first average. The sampling rate for the Air signals was 16000 Hz, so each resulting bin represented 0.9766 Hz. The 16385 independent values in the Air FFTs were thus compressed to a network input size of 32, each value thus covering a range of 500.0 Hz. Table 8.1.1.3.3-1 gives the final correspondence between bins and frequency ranges for the Air signals.

Free-Field Signal Set

Input	Frequency Range Covered			
1	243.16	to	259.77	Hz
2	259.77	to	275.39	Hz
3	275.39	to	291.02	Hz
4	291.02	to	306.64	Hz
5	306.64	to	322.27	Hz
6	322.27	to	337.89	Hz
7	337.89	to	353.52	Hz
8	353.52	to	369.14	Hz
9	369.14	to	384.77	Hz
10	384.77	to	400.39	Hz
11	400.39	to	416.02	Hz
12	416.02	to	431.64	Hz
13	431.64	to	447.27	Hz
14	447.27	to	462.89	Hz
15	462.89	to	478.52	Hz
16	478.52	to	494.14	Hz
17	494.14	to	509.77	Hz
18	509.77	to	525.39	Hz
19	525.39	to	541.02	Hz
20	541.02	to	556.64	Hz
21	556.64	to	572.27	Hz
22	572.27	to	587.89	Hz

Table 8.1.1.3.1-1 Free-Field Network Inputs in Frequency Domain

Bottom Signal Set

Input	Frequency Range Covered			
1	229.49	to	244.14	Hz
2	244.14	to	259.77	Hz
3	259.77	to	275.39	Hz
4	275.39	to	291.02	Hz
5	291.02	to	306.64	Hz
6	306.64	to	322.27	Hz
7	322.27	to	337.89	Hz
8	337.89	to	353.52	Hz
9	353.52	to	369.14	Hz
10	369.14	to	384.77	Hz
11	384.77	to	400.39	Hz
12	400.39	to	416.02	Hz
13	416.02	to	431.64	Hz
14	431.64	to	447.27	Hz
15	447.27	to	462.89	Hz
16	462.89	to	478.52	Hz
17	478.52	to	494.14	Hz
18	494.14	to	509.77	Hz
19	509.77	to	525.39	Hz
20	525.39	to	541.02	Hz
21	541.02	to	556.64	Hz
22	556.64	to	572.27	Hz
23	572.27	to	587.89	Hz

Table 8.1.1.3.2-1 Bottom Network Inputs in Frequency Domain

Air Signal Set

Input	Frequency Range Covered		
1	Offset + 0	to	500 Hz
2	500	to	1000 Hz
3	1000	to	1500 Hz
4	1500	to	2000 Hz
5	2000	to	2500 Hz
6	2500	to	3000 Hz
7	3000	to	3500 Hz
8	3500	to	4000 Hz
9	4000	to	4500 Hz
10	4500	to	5000 Hz
11	5000	to	5500 Hz
12	5500	to	6000 Hz
13	6000	to	6500 Hz
14	6500	to	7000 Hz
15	7000	to	7500 Hz
16	7500	to	8000 Hz
17	8000	to	8500 Hz
18	8500	to	9000 Hz
19	9000	to	9500 Hz
20	9500	to	10000 Hz
21	10000	to	10500 Hz
22	10500	to	11000 Hz
23	11000	to	11500 Hz
24	11500	to	12000 Hz
25	12000	to	12500 Hz
26	12500	to	13000 Hz
27	13000	to	13500 Hz
28	13500	to	14000 Hz
29	14000	to	14500 Hz
30	14500	to	15000 Hz
31	15000	to	15500 Hz
32	15500	to	16000 Hz

Figure 8.1.1.3.3-1 Air Network Inputs in Frequency Domain

8.1.1.3.4 Frequency Bin Definition

After the filtering performed on the various signals, the frequency domain inputs created from the Free-field, Bottom and Air signals each represented bands of width 15.6 kHz, 15.6 kHz and 500 Hz, respectively. For convenience, the frequency content of a particular input will be referred to by the lower bound of its range, with the true range of the band implied. For example, the Air signals were unfiltered, so input I1 in the Air frequency domain signals covers the frequencies 0 - 500 Hz. For brevity, in the context of discussion it would be said simply that input I1 corresponds to 0 Hz. Similarly, the statement that in the Bottom frequency domain, input I7 corresponds to 323 kHz really means that I7 corresponds to the range starting at 323 kHz, and continuing for another 15.6 kHz. In round figures, this is the range 323 - 339 kHz.

8.1.2 Network Training Using Clean Signals

Pilot studies were conducted to determine the values for the various adjustable network parameters, such as the learning rate. The values of these parameters, shown in Table 8.1.2-1, were fixed and common to all network runs. The number of input nodes for the networks varied with the signal condition, and are shown in Table 8.1.2-2.

Network Parameter	Setting Used
Learning Rule	Backpropagation - delta rule
Training	With Validation Set
Learning Rate	0.1
Momentum	0.5
Training Cycles	20,000
Input Noise	None
Validation Interval	10 Cycles

Table 8.1.2-1 Network Parameters

The number of hidden nodes was varied as an independent variable to evaluate the effect on the solution. For each condition, networks with hidden layers of 0 (a two layer network), 2, 4, and 6 hidden nodes were trained. The pilot studies indicated that the number of hidden nodes had a large effect on the network's ability to learn the patterns under consideration. The influence of the hidden nodes is studied in more detail in this experiment.

Signal Condition	Frequency Domain	Time Domain
Free-Field	22	25
Bottom	23	43
Air	32	32

Table 8.1.2-2 Number of Input Nodes for Frequency and Time Domains

There were always 7 output nodes, by which each network indicated its classification of the input signal by parameter. The output nodes and their corresponding parameters and the classes they represent are listed in Table 8.1.2-3.

Output Node	Parameter	Class Identified
1	Material	Brass
2	Material	Steel
3	Thickness	10%
4	Thickness	5%
5	Angle/Striker	0°/Metal
6	Angle/Striker	45°/Plastic
7	Angle/Striker	90°/Wood

Table 8.1.2-3 Output Node Description

Each output node had a target value of 0 or 1, which indicated the class to which the applied signal input belonged. A one on an output node indicated that the signal was of the corresponding class.

The sigmoid squashing function was always used as the transfer function for both the hidden and output layers.

A total of 72 neural networks were trained, 36 for the frequency domain signals and 36 for the time domain signals. The breakdown of the 36 runs is the same for each of the domains. There were 12 runs for each of the 3 signal conditions (Free-field, Bottom, and Air), and there were 3 runs for each for the 4 different hidden node possibilities (0, 2, 4, and 6). Runs with the same number of hidden nodes were differentiated by selecting a different random seed for initializing the weights, thereby starting the networks in a different position in the weight space. A summary of the number of neural networks that were run is shown in Table 8.1.2-4. The table is identical for both the frequency and time domains.

Network Configurations

Signal Condition	Hidden nodes	Number of Runs
Air	0	3
	2	3
	4	3
	6	3
Bottom	0	3
	2	3
	4	3
	6	3
Free-Field	0	3
	2	3
	4	3
	6	3
		36

Table 8.1.2-4 Neural Networks Run for Three Signal Conditions

All of the runs were performed on a SUN SparcStation, with a neural network program developed by ARD. Training was conducted for 20,000 cycles for all networks, with a cycle equaling one pass through the entire training set. Every ten cycles the validation set was presented to the network and the mean squared error was calculated. If the mean squared error was lower than all previous mean squared errors calculated for the validation set, the current weight matrix was maintained as the "best weights." At the end of the 20,000 cycles the "best weights" were captured for use in the analysis of network performance.

8.1.3 Clean-Trained Networks Tested with Clean Signals

Two benchmarks were used to determine the neural networks performance on the validation set, mean squared error and percent correct. These two benchmarks are defined as follows.

Mean squared error minus the sum of the outputs minus the targets squared, for each of the 96 validation patterns divided by 96 (the number of validation patterns).

Percent correct - Percent correct was broken down into four categories
% correct Material
% correct Thickness
% correct Angle (Striker for the Air signals)
% correct Overall (all three parameters correct)

Percent correct refers to the proportion of the validation patterns that the network was able to classify correctly. A simple algorithm was used to calculate the percent correct. For example, to determine the percent correct for Angle the following procedure was used. There are three output nodes that represented Angle (0°, 45°, and 90°). One of the target outputs for these three nodes was always 1 and the others were always zero. If the value of the output node with a target value of one is greater than the output values from the other two output nodes, then this pattern is counted as correct for Angle. Similar calculations are done for Material and Thickness. Percent correct overall is the percent of the patterns that were simultaneously correct (as defined above) for Material, Thickness, and Angle.

8.1.3.1 Frequency Domain Results

The results of the frequency domain neural networks are summarized in Tables 8.1.3.1-1 through 8.1.3.1-3 and Figures 8.1.3.1-1 through 8.1.3.1-3. The tables show the percent correct and mean squared error for each of the 36 frequency domain runs for each signal condition along with averages across random seed. The figures show the same data for the single best network at each number of hidden nodes. Some networks had perfect performance (100% correct for the Overall condition) for both Free-field and Bottom signals. Thus it appears that the neural networks are well suited for these signals in the frequency domain. Performance on the Air signals was also very good (97%), but it never reached the 100% levels achieved by the Free-field and Bottom neural networks.

The performance of the Air signal neural networks was very high except for the 2 hidden node case where the average percent correct (all) was only 69.3%. This contrasts with the 0, 4, and 6 hidden node conditions for which average percent correct (all) is near 100%. The mean squared error follows a similar pattern with 2 hidden nodes being the worst and the other conditions having a much lower error. The performance of the Bottom signal neural networks was 100% correct for the 0, 4, and 6 hidden node conditions. The percent correct was only about 70% for the 2 hidden

Hidden Nodes	Parameter	Seed 1	Seed 2	Seed 3	Average
0	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.002	0.002	0.002	0.002
2	M	99.0	95.8	96.9	97.2
	T	97.9	100.0	97.9	98.6
	A	78.1	74.0	65.63	72.6
	All	75.0	74.0	61.5	70.1
	MSE	0.471	0.504	0.544	0.506
4	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	91.7	97.2
	All	100.0	100.0	91.7	97.2
	MSE	0.001	0.000	0.086	0.029
6	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.000	0.000	0.000	0.000

Average Performance Across Seeds

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	97.2	100.0	100.0
T	100.0	98.6	100.0	100.0
A	100.0	72.6	97.2	100.0
All	100.0	70.1	97.2	100.0
MSE	0.002	0.506	0.029	0.000

Best Network Performance

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	99.0	100.0	100.0
T	100.0	97.9	100.0	100.0
A	100.0	78.1	100.0	100.0
All	100.0	75.0	100.0	100.0
MSE	0.002	0.47	0.00	0.000

Table 8.1.3.1-1 Free-Field Frequency Domain Network Performance

Hidden Nodes	Parameter	Seed 1	Seed 2	Seed 3	Average
0	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.007	0.007	0.007	0.007
2	M	78.1	83.3	75.0	78.8
	T	66.7	100.0	100.0	88.9
	A	100.0	100.0	100.0	100.0
	All	54.2	83.3	75.0	70.8
	MSE	0.664	0.501	0.546	0.570
4	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.000	0.000	0.000	0.000
6	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.000	0.000	0.000	0.000

Average Performance Across Seeds

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	78.8	100.0	100.0
T	100.0	88.9	100.0	100.0
A	100.0	100.0	100.0	100.0
All	100.0	70.8	100.0	100.0
MSE	0.007	0.570	0.000	0.000

Best Network Performance

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	83.3	100.0	100.0
T	100.0	100.0	100.0	100.0
A	100.0	100.0	100.0	100.0
All	100.0	83.3	100.0	100.0
MSE	0.007	0.50	0.00	0.00

Table 8.1.3.1-2 Bottom Frequency Domain Network Performance

Hidden Nodes	Parameter	Seed 1	Seed 2	Seed 3	Average
0	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	S	96.9	96.9	96.9	96.9
	All	96.9	96.9	96.9	96.9
	MSE	0.087	0.089	0.088	0.088
2	M	100.0	100.0	100.0	100.0
	T	96.9	97.9	99.0	97.4
	S	63.5	79.2	75.0	71.4
	All	61.5	77.1	74.0	69.3
	MSE	0.544	0.447	0.468	0.496
4	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	S	96.9	97.9	97.9	97.6
	All	96.9	97.9	97.9	97.6
	MSE	0.042	0.032	0.036	0.037
6	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	S	97.9	97.9	97.9	97.9
	All	97.9	97.9	97.9	97.9
	MSE	0.035	0.036	0.046	0.039

Average Performance Across Seeds

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	100.0	100.0	100.0
T	100.0	97.4	100.0	100.0
S	96.9	71.4	97.6	97.9
All	96.9	69.3	97.6	97.9
MSE	0.088	0.496	0.037	0.039

Best Network Performance

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	100.0	100.0	100.0
T	100.0	97.9	100.0	100.0
S	96.9	79.2	97.9	97.9
All	96.9	77.1	97.9	97.9
MSE	0.09	0.45	0.03	0.03

Table 8.1.3.1-3 Air Frequency Domain Network Performance

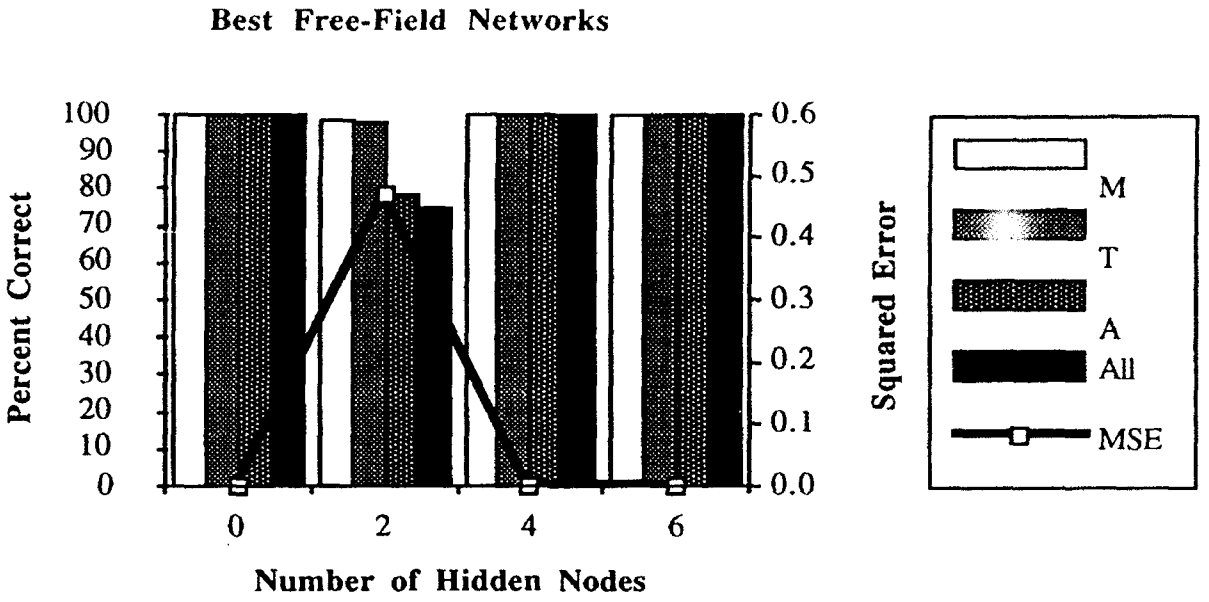


Figure 8.1.3.1-1 Performance for Best Free-Field Frequency Domain Network

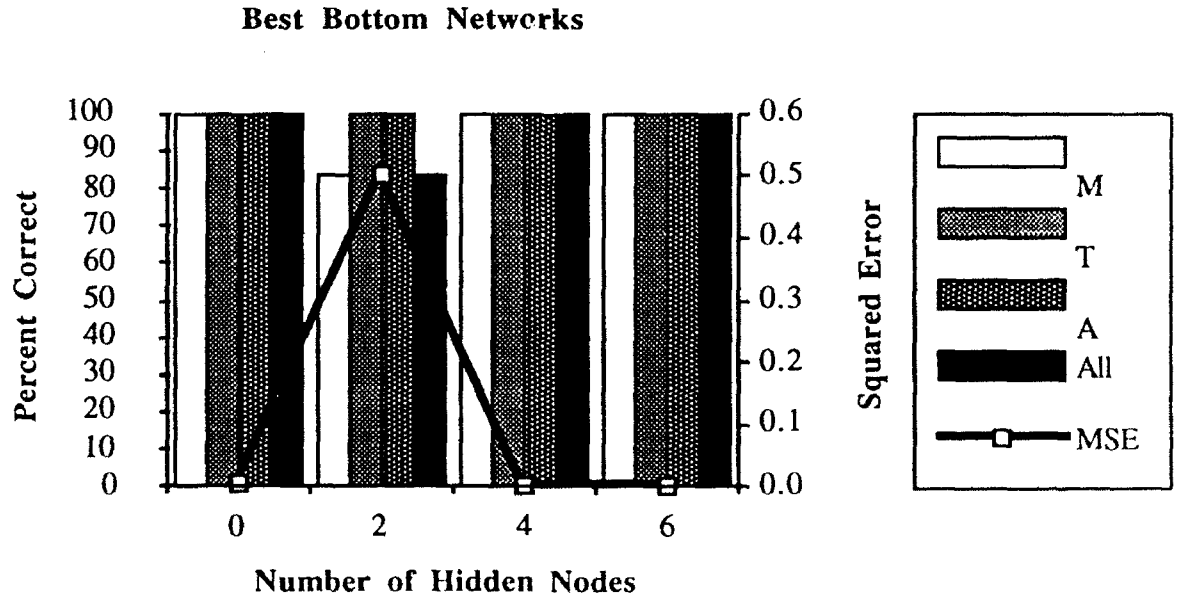


Figure 8.1.3.1-2 Performance for Best Bottom Frequency Domain Network

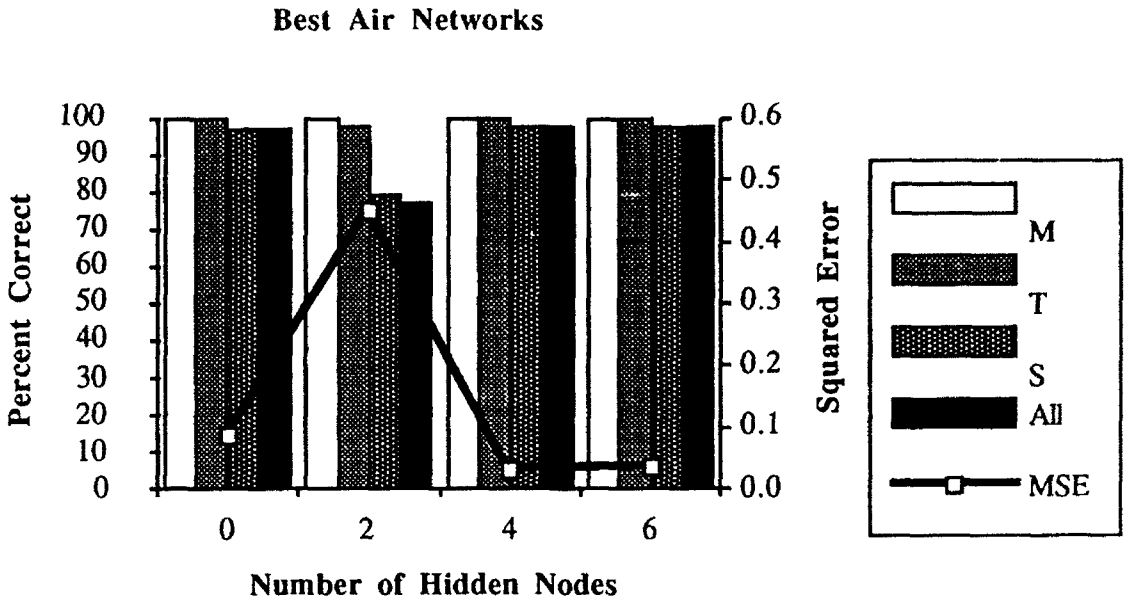


Figure 8.1.3.1-3 Performance for Best Air Frequency Domain Network

node condition. Similarly, the performance of the Free-field signal neural networks was also very good except for the 2 hidden node condition.

Since the networks without hidden layers successfully classified the signals, it is clear that the problem can be accomplished without nonlinear elements. The two hidden node networks had the benefit of nonlinear elements, yet were generally less capable of the classification tasks. Presumably the two hidden node networks lacked enough network connections on which to encode a sufficient solution. For instance, an Air network without a hidden layer had 32 * 7 = 224 connections. Given two hidden nodes, the network had only 78 connections. The advantages of a nonlinear transformation could not overcome the relative lack of connections.

8.1.3.2 Time Domain Results

The results of the time domain neural networks are summarized in Tables 8.1.3.2-1 through 8.1.3.2-3 and Figures 8.1.3.2-1 through 8.1.3.2-3. The tables show the percent correct and mean squared error for each of the 36 time domain runs for one type of signal along with averages across

Hidden Nodes	Parameter	Seed 1	Seed 2	Seed 3	Average
0	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.122	0.122	0.122	0.122
2	M	83.3	64.6	58.3	68.7
	T	66.7	76.0	75.0	72.6
	A	96.9	100.0	99.0	98.6
	All	46.9	54.2	40.6	47.2
	MSE	0.789	0.691	0.824	0.768
4	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.001	0.001	0.001	0.001
6	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.000	0.000	0.000	0.000

Average Performance Across Seeds

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	68.7	100.0	100.0
T	100.0	72.6	100.0	100.0
A	100.0	98.6	100.0	100.0
All	100.0	47.2	100.0	100.0
MSE	0.122	0.768	0.001	0.000

Best Network Performance

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	64.6	100.0	100.0
T	100.0	76.0	100.0	100.0
A	100.0	100.0	100.0	100.0
All	100.0	54.2	100.0	100.0
MSE	0.122	0.691	0.001	0.000

Table 8.1.3.2-1 Free-Field Time Domain Network Performance

Hidden Nodes	Parameter	Seed 1	Seed 2	Seed 3	Average
0	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.001	0.001	0.001	0.001
2	M	51.0	77.1	89.6	72.6
	T	66.7	50.0	89.6	68.8
	A	100.0	100.0	100.0	100.0
	All	33.3	38.5	79.2	50.3
	MSE	0.665	0.687	0.560	0.637
4	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.000	0.000	0.000	0.000
6	M	100.0	100.0	100.0	100.0
	T	100.0	100.0	100.0	100.0
	A	100.0	100.0	100.0	100.0
	All	100.0	100.0	100.0	100.0
	MSE	0.000	0.000	0.000	0.000

Average Performance Across Seeds

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	72.6	100.0	100.0
T	100.0	68.8	100.0	100.0
A	100.0	100.0	100.0	100.0
All	100.0	50.3	100.0	100.0
MSE	0.001	0.637	0.000	0.000

Best Network Performance

	Hidden Nodes			
Parameter	0	2	4	6
M	100.0	89.6	100.0	100.0
T	100.0	89.6	100.0	100.0
A	100.0	100.0	100.0	100.0
All	100.0	79.2	100.0	100.0
MSE	0.001	0.560	0.000	0.000

Table 8.1.3.2-2 Bottom Time Domain Network Performance

Hidden Nodes	Parameter	Seed 1	Seed 2	Seed 3	Average
0	M	99.0	99.0	99.0	99.0
	T	100.0	100.0	100.0	100.0
	S	71.9	71.9	75.0	72.9
	All	71.9	71.9	75.0	72.9
	MSE	0.503	0.502	0.501	0.502
2	M	99.0	74.0	100.0	86.5
	T	99.0	100.0	95.8	99.5
	S	60.4	38.5	41.7	49.5
	All	58.3	25.0	40.6	41.7
	MSE	0.656	1.043	0.713	0.850
4	M	96.9	99.0	99.0	98.3
	T	95.8	99.0	97.9	97.6
	S	41.7	84.4	83.3	69.8
	All	36.5	83.3	83.3	67.7
	MSE	0.723	0.255	0.262	0.413
6	M	100.0	100.0	100.0	100.0
	T	97.9	96.9	96.9	97.2
	S	89.6	85.4	87.5	87.5
	All	87.5	82.3	84.4	84.7
	MSE	0.236	0.292	0.267	0.265

Average Performance Across Seeds

	Hidden Nodes			
Parameter	0	2	4	6
M	99.0	86.5	98.3	100.0
T	100.0	99.5	97.6	97.2
S	72.9	49.5	69.8	87.5
All	72.9	41.7	67.7	84.7
MSE	0.502	0.850	0.413	0.265

Best Network Performance

	Hidden Nodes			
Parameter	0	2	4	6
M	99.0	99.0	99.0	100.0
T	100.0	99.0	99.0	97.9
S	75.0	60.4	84.4	89.6
All	75.0	58.3	83.3	87.5
MSE	0.501	0.656	0.255	0.236

Table 8.1.3.2-3 Air Time Domain Network Performance

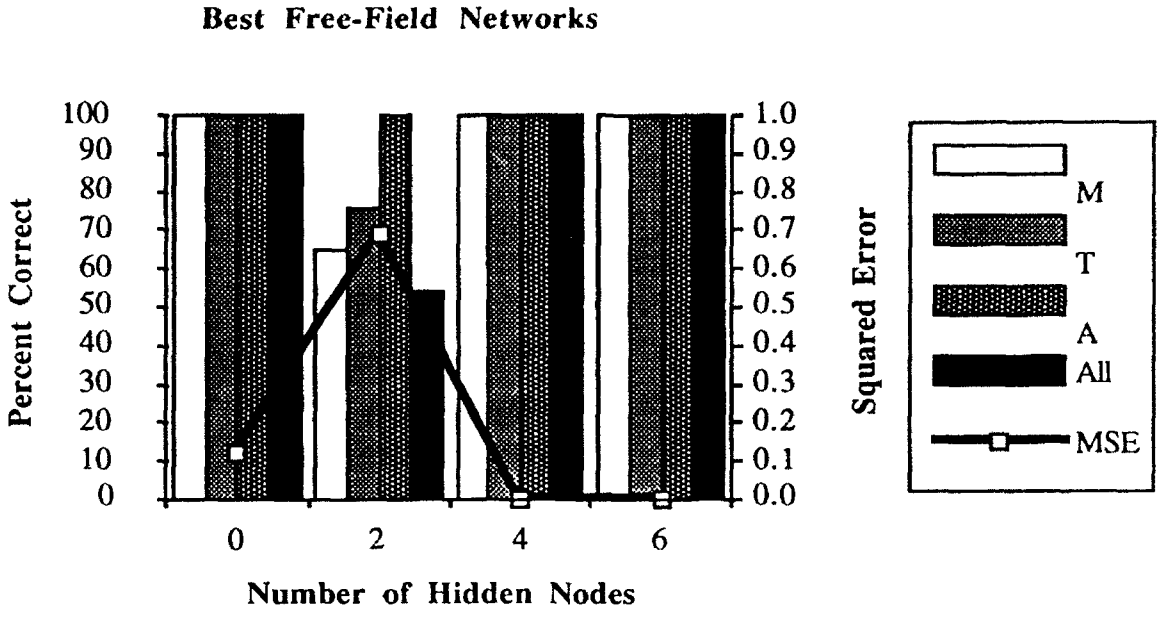


Figure 8.1.3.2-1 Performance for Best Free-Field Time Domain Network

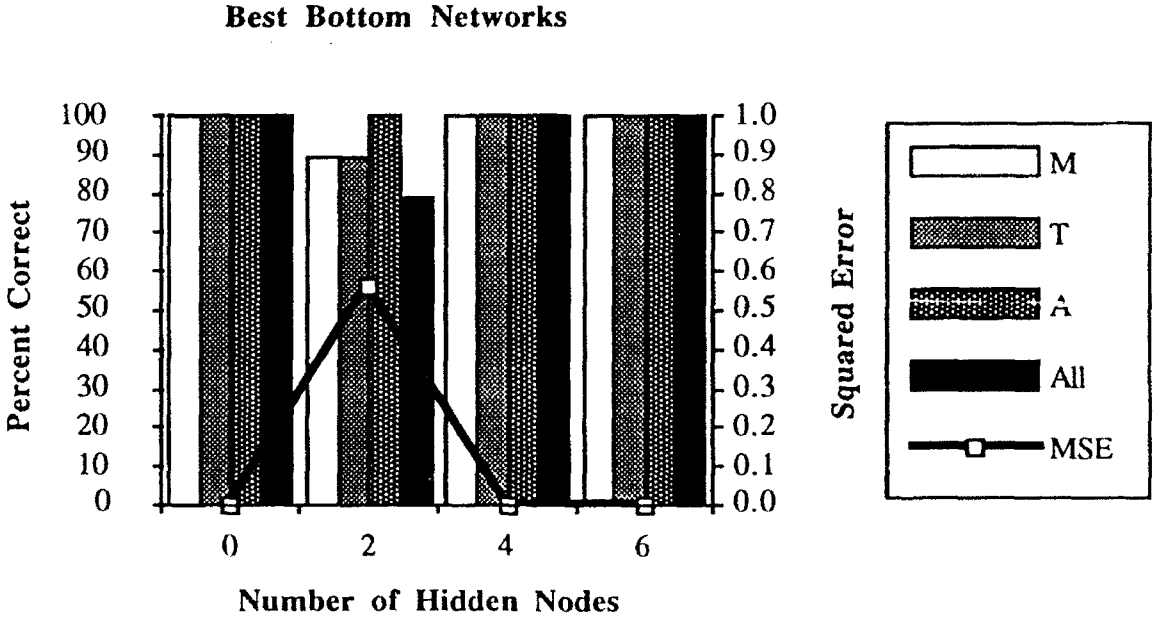


Figure 8.1.3.2-2 Performance for Best Bottom Time Domain Network

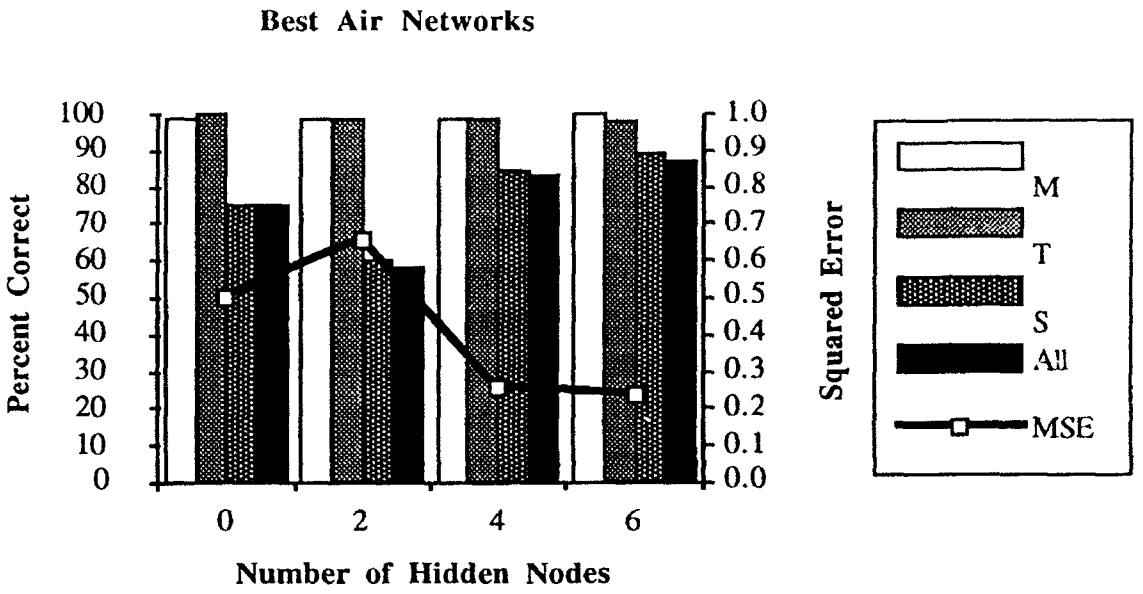


Figure 8.1.3.2-3 Performance for Best Air Time Domain Network

random seeds. The figures show the same data for the single best network at each number of hidden nodes.

The Free-field and Bottom networks had very little problem reaching the 100% correct level. Just as in the frequency domain the neural networks did not have any problem making perfect classifications using the time domain representation. The performance on the Air signals is somewhat worse for the time domain signals. The highest level of performance was for the 6 hidden node condition where the performance reached 87%.

8.1.3.3 Discussion of Performance for Clean-Trained Networks

The performance of the Free-field neural networks trained with signals with no added noise was always at 100% correct except for the 2 hidden node condition. The 2 hidden node neural networks only reached 47% average overall correct. Time and frequency domain input networks performed similarly except for the 2 hidden node cases, in which frequency domain input was preferable.

The performance of the Bottom neural networks was always at 100% correct except for the 2 hidden node condition. In the 2 hidden node condition, the average percent correct Overall just

reached 50% in the time domain case. Signal representation was not a large factor in performance aside from the two hidden node cases. When the Bottom networks were faced with the challenging condition of having two hidden nodes, Angle performance remained at 100% while Material and Thickness performances fell. This effect was consistent across signal representation (frequency and time domains). Angle is apparently easier for these networks to classify. This is easy to understand for the 90° signals in the time domain, since they have a significantly different envelope than the other angles. The networks, however, could also tell 0° from 45° signals with only 2 hidden nodes, and could do so using frequency domain input as well.

The performance of Air neural networks varied greatly across the different number of hidden nodes. In general, the 6 hidden node condition had the best performance with 0 and 4 hidden nodes very close in performance and the 2 hidden node case well below the others. However, the best performance on the time domain Air signals was only about 85% correct. This compares to almost 98% correct on the frequency domain signals. This lower performance was primarily due to the decrease in performance on Striker when time domain input was employed. Striker was the most difficult parameter for every case of signal representation and number of hidden nodes.

It is interesting to note that the networks trained to classify Air signals as time domain input performed worse than did networks trained to classify Bottom and Free-field signals. This is the opposite effect observed in the human results, in which subjects found the Air signals easier to classify.

8.1.4 Clean-Trained Networks Tested with Noisy Signals

The performance of the original networks was evaluated by several criteria. The most natural and immediate was their ability to classify the original ninety-six test signals. The results of these tests were described above. The resilience of the networks to the presence of background noise is a more informative measure, for two reasons. First, a network which is tolerant of noise will operate under a larger range of signal conditions, which makes it more useful than one which can only classify clean signals. The lower the signal-to-noise ratio that a network can tolerate, the more robust a classifier it is. Second, testing the networks on moderately noisy signals provides information about the generality of the algorithms the networks have developed. In principle, a network which has learned to classify the signals correctly on the basis of general traits of the signal classes would be expected to classify correctly an infinite number of examples of any given

signal class. On the other hand, if the network classifies the signals on the basis of artifacts peculiar to the training or testing sets, it may incorrectly classify signals which are even slightly different from the original ninety-six test signals. By adding sequences of noise to the original ninety-six test signals it was possible to create many new test signals which resembled the original signals, but did not match them exactly, and thus test the generality of the networks' algorithms.

The pseudo-random noise generated for this purpose was normally distributed about a mean of zero, and hence completely characterized by its standard deviation (see Figure 8.1.4-1). Each of the ninety-six signals in the original test set was used to generate twenty different noisy signals in each new test set. This redundancy was included to reduce any effects arising spuriously from the characteristics of particular pseudo-random number sequences. The seed used to start the pseudo-random number sequences was also varied throughout the tests.

Multiple test sets were created whose standard deviation spanned the range (0.0, 2.0). By testing the networks on each of these new test sets the resilience of the networks to the presence of noise was investigated. Results for Bot4H(1)F are shown in Figure 8.1.4-2, in which the root mean squared (RMS) error and percentage of correct classifications are plotted as a function of the standard deviation of the noise used to create the test set. As might be expected, with increasing noise the network's performance deteriorated from the level achieved by the networks on the original (clean) test set. This behavior was the same for every network tested; as noise increased, the percent of correct classifications dropped, approaching a plateau value between eight and ten percent. Remember that the odds of randomly classifying a signal correctly are one in twelve, or 8.33%. The exact rate of deterioration of performance depended on the domain, signal set and number of hidden nodes.

The results for the twenty-four best performing backpropagation networks trained with clean signals are summarized in Table 8.1.4-1. The first two columns list the percent of correct classifications and the RMS error of each network when tested on clean signals. On a graph such as Figure 8.1.4-2, these two quantities correspond to the y-intercepts of the percent correct and RMS error, respectively. The third column shows the noise test 30% point, namely, the standard deviation of added noise at which the given network's performance dropped below 30% correct. The noise test 30% point is also shown graphically in Figure 8.1.4-2 for the network Bot4H(1)F. This latter value, combined with the percent correct, gives some indication of how rapidly the performance falls to its final value. For example, among Air networks in the frequency domain,

Figure 1(a)

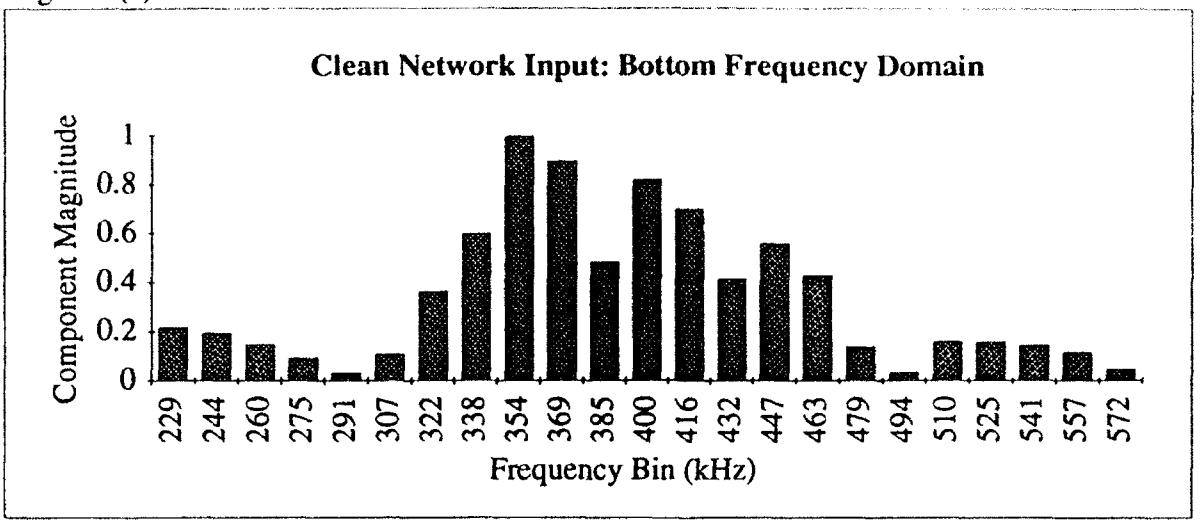


Figure 1(b)

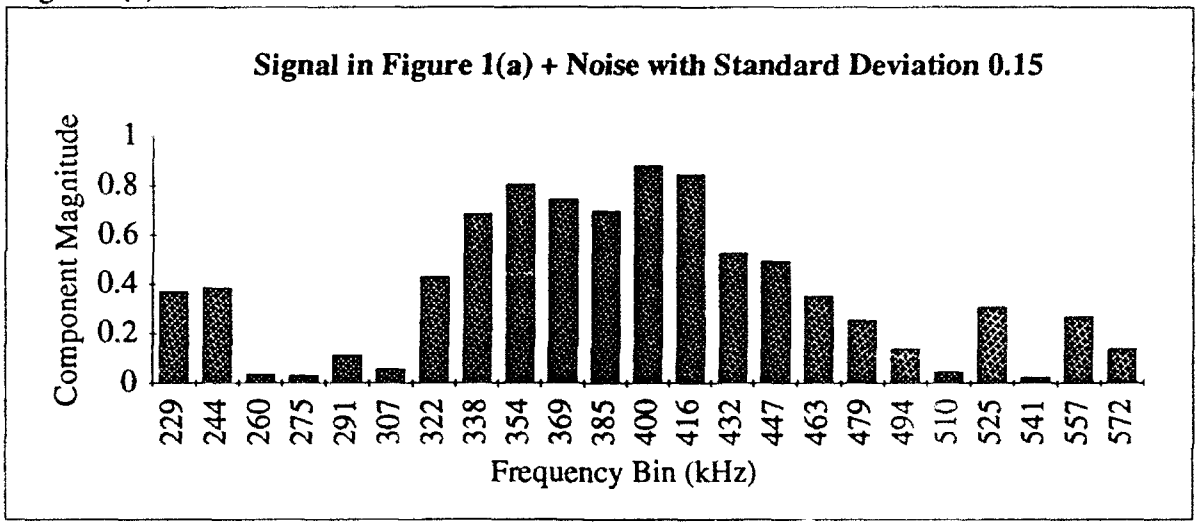


Figure 1(c)

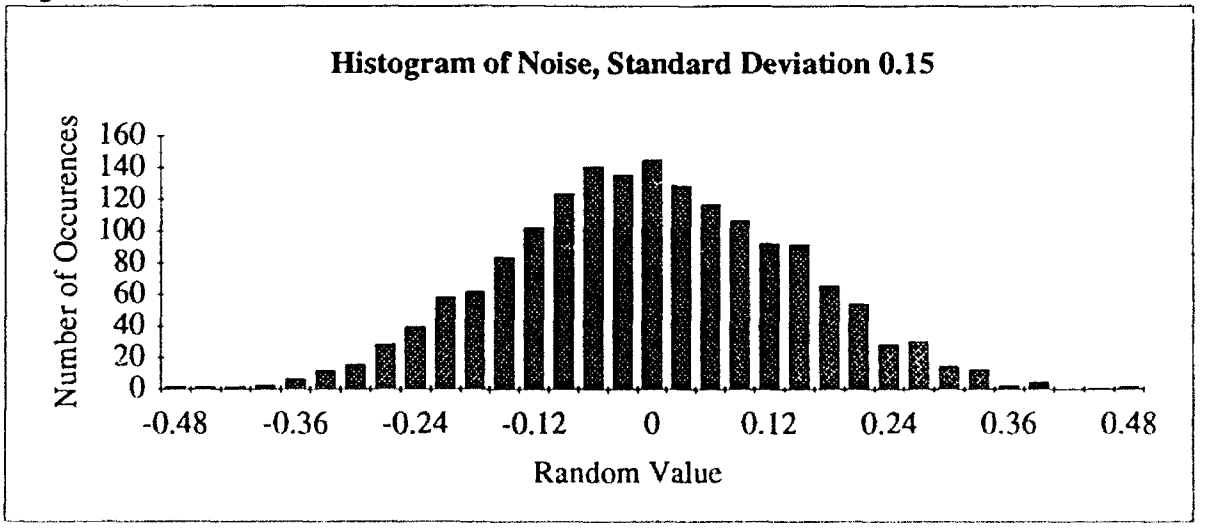


Figure 8.1.4-1 Bottom Network Frequency Domain Input and Added Pseudo-Random Noise

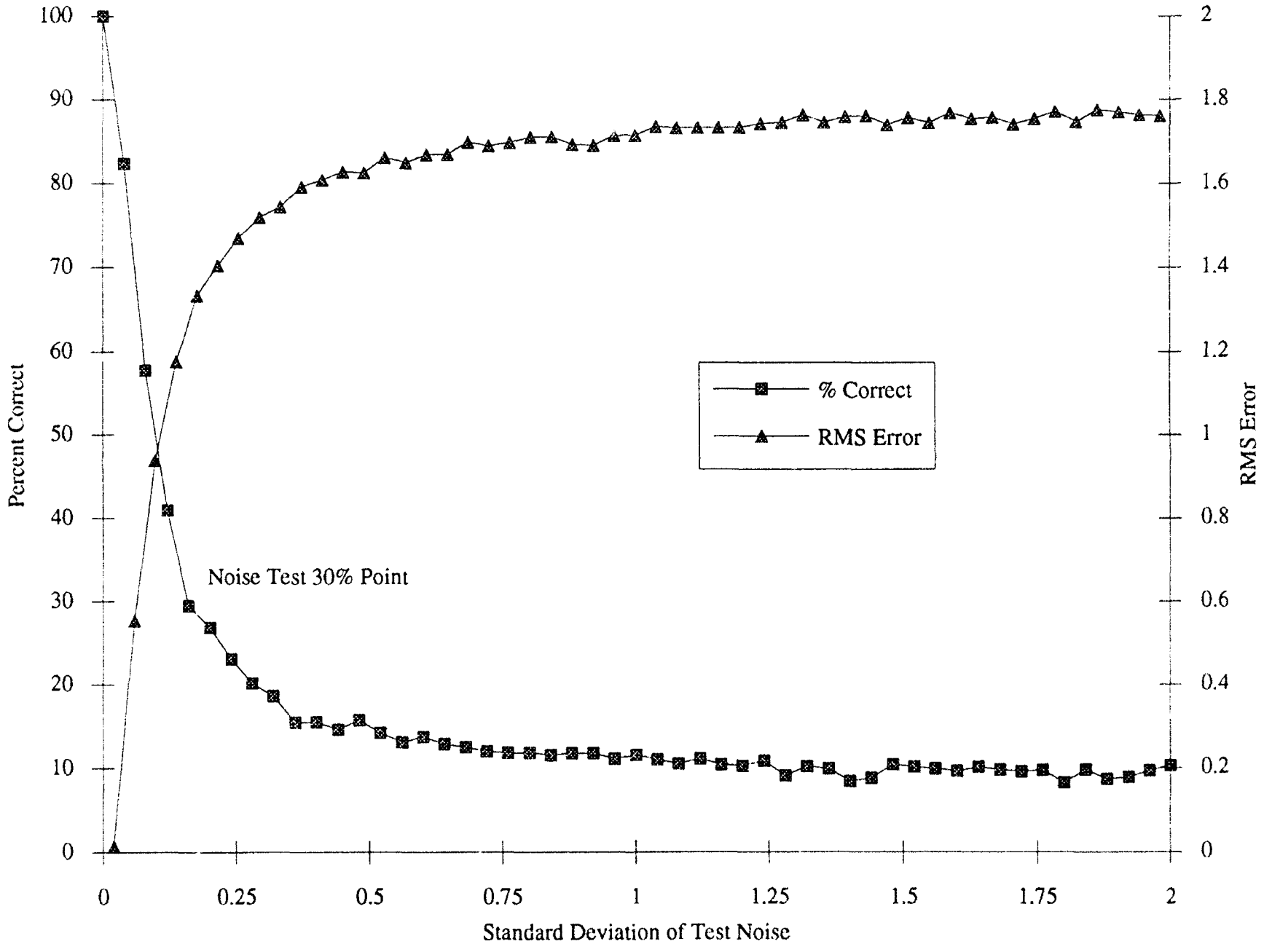


Figure 8.1.4-2 Bot4H(1)F Tested on Noisy Signals

Table 1(a): Frequency Domain

Network	% Correct Clean Signals	RMS Error Clean Signals	Noise Test 30% Point
Free 0h-3	100.00	0.04	0.26
Free 2h-1	75.00	0.69	0.14
Free 4h-2	100.00	0.02	0.25
Free 6h-3	100.00	0.01	0.23
Bot 0h-2	100.00	0.08	0.19
Bot 2h-2	83.33	0.71	0.08
Bot 4h-1	100.00	0.01	0.16
Bot 6h-1	100.00	0.01	0.21
Air 0h-1	96.88	0.30	0.52
Air 2h-2	77.08	0.67	0.10
Air 4h-2	97.92	0.18	0.50
Air 6h-1	97.92	0.19	0.52

Table 1(b): Time Domain

Network	% Correct Clean Signals	RMS Error Clean Signals	Noise Test 30% Point
Free 0h-1	100.00	0.35	0.18
Free 2h-2	54.17	0.83	0.14
Free 4h-2	100.00	0.02	0.19
Free 6h-1	100.00	0.01	0.28
Bot 0h-2	100.00	0.03	0.41
Bot 2h-3	79.17	0.83	0.15
Bot 4h-2	100.00	0.01	0.33
Bot 6h-1	100.00	0.01	0.36
Air 0h-3	75.00	0.71	0.04
Air 2h-1	58.33	0.81	0.03
Air 4h-2	83.33	0.50	0.06
Air 6h-2	82.29	0.54	0.09

Table 8.1.4-1 Clean-Trained Networks' Performance Summary

the lowest noise test 30% point is 0.10, occurring for the two hidden node network. This is much smaller than the noise test 30% points of the zero, four and six node Air frequency domain networks (0.52, 0.50 and 0.52, respectively). However, the two hidden node network also achieved only 77.08 % correct on clean signals - significantly less than the percent correct for the other Air frequency domain networks. Therefore, while it is true that the two node network falls from its best performance faster than the others under the influence of noise, the difference is not as extreme as the noise test 30% point alone would lead one to believe.

With the exception of the two node networks, the networks performed fairly well (more than fifty percent correct), provided the standard deviation of the noise remained less than or equal to about 0.1. It should be borne in mind that in both the frequency and time domains the original signals presented to the network were normalized to have values between 0.0 and 1.0. Noise of standard deviation 0.1 therefore implies a distribution of noise whose width is 10 % of the signal's maximum value. From this perspective, the clean-trained networks show some amount of learning generality in their performance.

8.1.5 Networks Trained using Noisy Signals

In the experiments discussed above, noisy signals were used only for testing the networks, and not for training or validation. Perhaps of more interest is the question of what influence, if any, the addition of noise to the signals during training has on the performance. It was thought that the addition of some noise during training would in effect enlarge the training set, and obscure small, random variations in the signals, forcing the network to learn a more general solution. A network trained in this way might tolerate larger variations in the test set, performing better on noisy signals. On the other hand, if too much training noise were added, the networks might not learn to detect features in the training signals, and consequently would perform very badly, even on clean test signals. The level of training noise was therefore an important parameter to determine. A second issue was the choice of validation set, which is used to determine the " best" set of network weights. It was unclear whether clean signals, noisy signals, or some combination should be used. The first step then was to focus on a particular network to resolve these two questions, thereby standardizing the noise levels for the training and validation sets. The network chosen for these experiments was a 4 hidden node Bottom network using signals in their frequency domain form.

8.1.5.1 Noise Level for Training and Validation Sets

The first issue explored was the choice of validation set. Three networks were trained from identical initial conditions, with a noise level of 0.05, but validated with three different sets. One set was just the original (clean) test set. A second consisted of signals to which noise with standard deviation 0.05 had been added and a third contained a mix of the two. As with the noisy test sets described above, each clean signal was used multiple times to generate noisy validation signals. For ease of implementation, each signal was used only 5 times for validation sets, instead of 20. In the set containing the mixture, the clean signals were simply repeated five times to assure equal representation. The best weights chosen in each case were identical. Additional tests with training noise levels of 0.0 (clean training) and 0.10, and the validation sets described above, again failed to show any differences in best weight selection. Several additional variations in the validation set were then tried, including noise levels as high as 0.15, with no change in the set of best weights chosen. These results show that the choice of validation set did not influence the choice of best weights for the specific case of Bottom, four hidden node, frequency domain networks. Since the tests indicated no preference for a particular validation set, a standard procedure for creating validation sets in the other domains using different signal sets remained unclear. The standard procedure finally set was to use a mixture of clean signals and signals with noise of standard deviation 0.05, in equal quantity. The reason for this choice was simply that a result of tests on a single network was being generalized to determine a procedure for all the networks, and this mixture was thought to be the least "risky" in the event that the other domains were not identical in their responses to validation sets.

Once the validation set was standardized, the only remaining parameter to fix was the training noise level. Once again, a series of initially identical networks was trained, this time with training noise levels with standard deviations of 0.0, 0.03, 0.05, 0.07 and 0.10. The three curves in Figure 8.1.5.1-1(a) show an enhancement of the classification performance as the training noise level was increased from 0.0 (no training noise) through 0.05. The largest improvement over the control network (no training noise) was 16.62%, occurring when the levels of training and testing noise were 0.05 and 0.06, respectively. Improvement was most striking for test noise of standard deviation less than 0.24, but the effect was noticeable for values of test noise as high as 1.0. As the training noise level was increased beyond 0.05 to 0.1, however, the performance dropped quickly, particularly for values of test noise under 0.1 (see Figure 8.1.5.1-1(b)). The training

Figure 1(a)

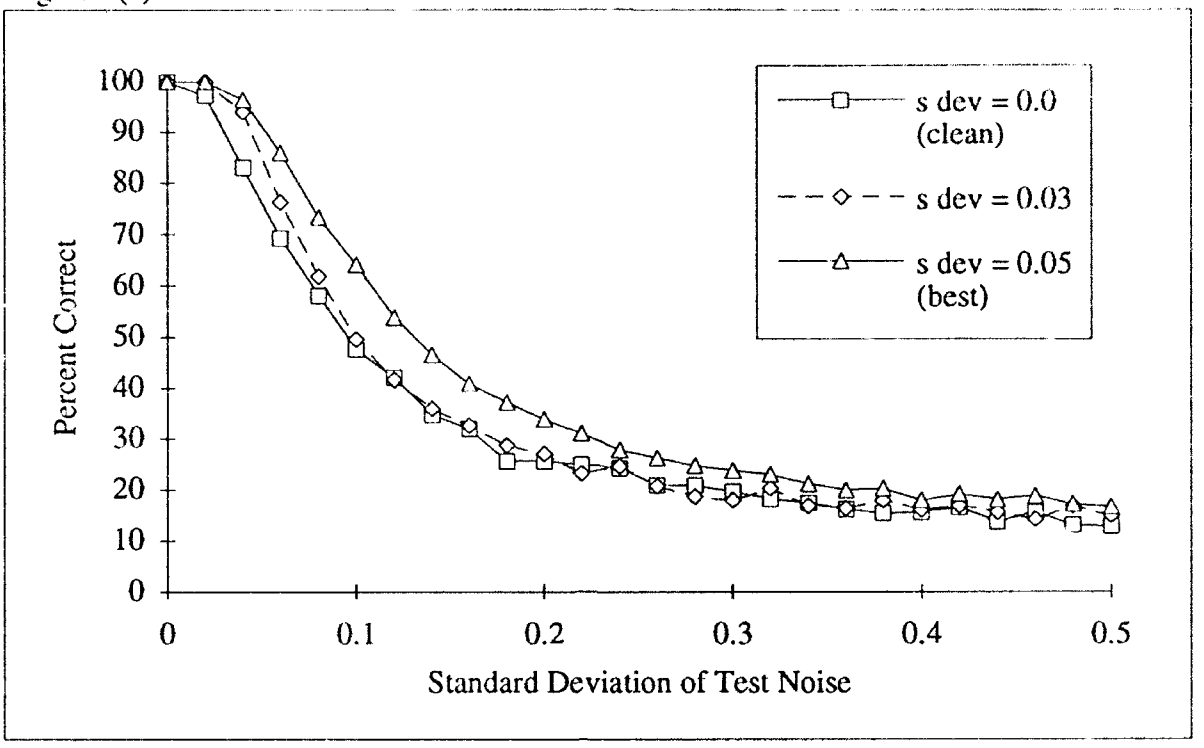


Figure 1(b)

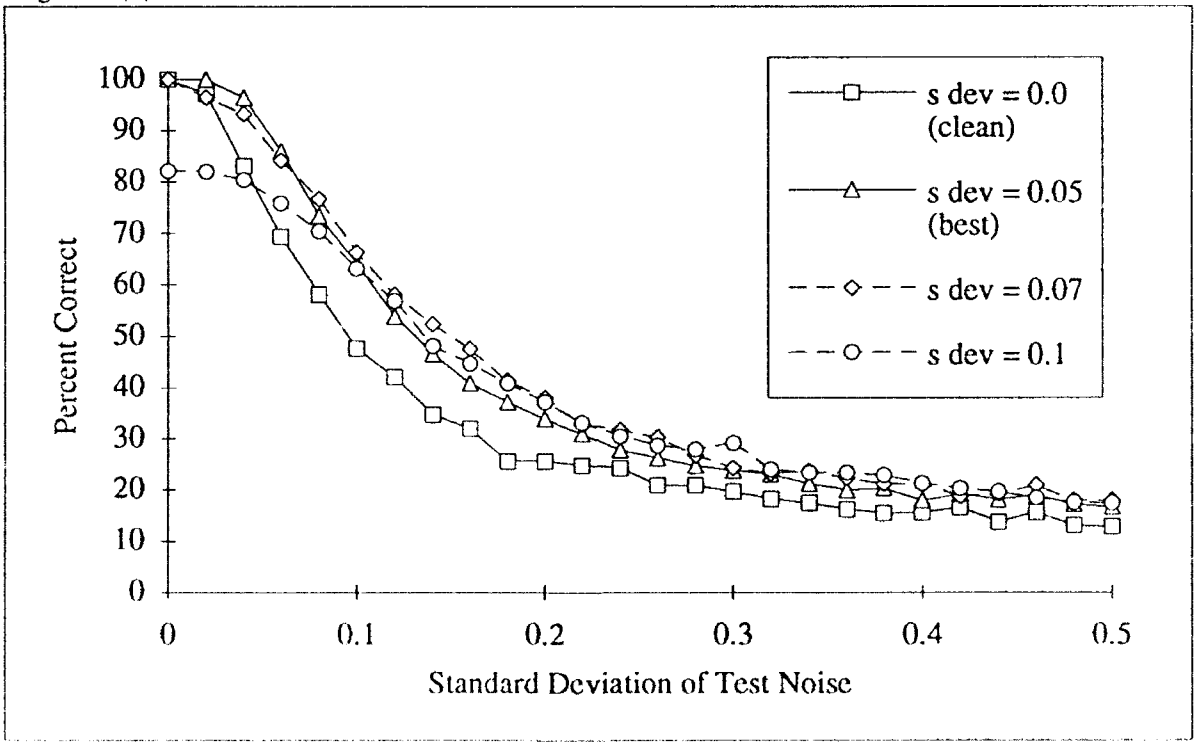


Figure 8.1.5.1-1 Effects on Network Performance as Noise Level on Training Data is Increased

noise level for all subsequent frequency domain networks was thus chosen to be 0.05, since this provided the most consistent enhancement over the broadest range of testing noise levels.

8.1.5.2 Training Regime

With all the parameters standardized, Bottom and Air networks in both frequency and time domains were retrained with noisy signals from the same initial conditions as the best performing clean-trained networks. For each domain and signal set, a validation set was created which contained equal portions of clean signals, and signals to which noise of standard deviation 0.05 had been added. The signals used in the training set had noise with a standard deviation of 0.05 added to them, and the networks were trained with all other network parameters (i.e. learning rate, number of cycles, etc.) identical to those used for the clean-trained networks. Among the clean-trained networks, those with 6 hidden nodes performed in all respects similarly to those with 4 hidden nodes. For this reason, only 0, 2, and 4 hidden node networks from each domain and signal set were retrained with noise. After these were trained, the networks were tested over a range of noise levels, in exactly the manner described above for clean-trained networks.

8.1.6 Noise-Trained Networks Tested with Clean and Noisy Signals

A typical result is shown in Figure 8.1.6-1, in which the percent of correct classifications is plotted for the networks Air2H(1)T and Air2H(1)TN. As was the case in the Bottom, 4 hidden node, frequency domain networks described above, the Air networks trained with noise show improved resilience to the presence of test signal noise. It is worth noticing that the clean-trained network, Air2H(1)T, classified clean signals (noise level 0.0) better than Air2H(1)TN. This was true of all the 2 hidden node networks, and several 0 and 4 hidden node networks as well. This may reflect a training noise level which is high enough to obscure clues essential to correct classification.

The results for the retrained networks are summarized in Table 8.1.6-1. The first column lists the name of each network. The next 3 columns show the same performance measures displayed for the clean-trained networks shown in Table 8.1.4-1. These are the percent of correct classifications and RMS error from tests on clean signals, and the noise test 30% point described above. The last column displays the Average Improvement of the networks trained with noise over those trained without noise. This value is the average difference per test point between the percent correct achieved by networks trained with and without noise, over the first 21 test levels. On a graph such

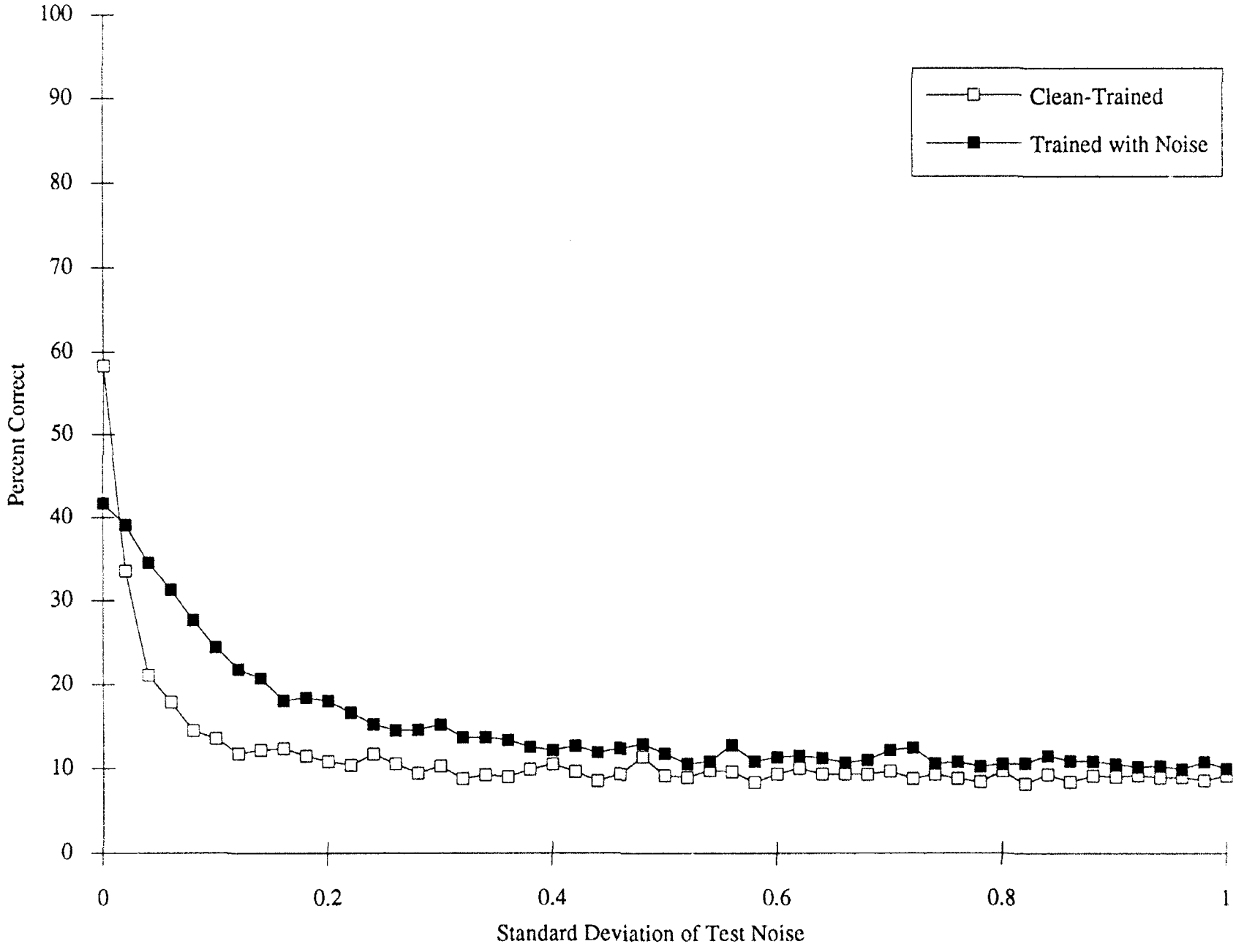


Figure 8.1.6-1 Comparison of Air2H(1) Performance When Trained with Clean and Noisy Input Data

Table 2(a): Frequency Domain

Network	% Correct Clean Signals	RMS Error Clean Signals	Noise Test 30% Point	Average Improvement Over Clean-Trained
Bot0H(2)FN	86.46	0.37	0.21	-0.02
Bot2H(2)FN	50.00	0.85	0.12	0.62
Bot4H(1)FN	100.00	0.11	0.23	7.56
Air0H(1)FN	95.83	0.33	0.52	3.14
Air2H(2)FN	70.83	0.71	0.32	18.45
Air4H(2)FN	97.92	0.21	0.40	-2.66

Table 2(b): Time Domain

Network	% Correct Clean Signals	RMS Error Clean Signals	Noise Test 30% Point	Average Improvement Over Clean-Trained
Bot0H(2)TN	100.00	0.03	0.39	-0.47
Bot2H(3)TN	66.67	0.78	0.26	7.83
Bot4H(2)TN	100.00	0.03	0.35	5.10
Air0H(3)TN	50.00	0.90	0.08	2.87
Air2H(1)TN	41.67	0.90	0.07	5.73
Air4H(2)TN	68.75	0.64	0.23	15.90

Table 8.1.6-1 Noise-Trained Networks' Performance Summary

as Figure 8.1.6-1, this value is proportional to the area between the curves, over the range (0.0, 0.4) in test noise levels. A positive value indicates that the network performance improved with training using noisy signals, a negative value shows the reverse.

For most of the networks, training noise had either no effect on performance, or a beneficial one. There are, however, three negative values appearing in Table 8.1.6-1 which deserve some scrutiny. The negative values occurring for Bot0H(2)FN and Bot0H(2)TN are negligible. Comparisons of these two networks to their clean-trained antecedents are shown in Figure 8.1.6-2. It is clear from the graphs that the clean-trained networks' performance is essentially unchanged by training with noise, except in with low noise levels where the standard deviation less than 0.04. This can be seen in Figure 8.1.6-2(a) for the Bottom, 4 hidden node, frequency domain network. By contrast, the Average Improvement of -2.66 achieved by Air4H(2)FN indicates that the performance was actually significantly worse for the network trained with noise (see Figure 8.1.6-3). These few results are contrary to the general trend followed by all the other networks trained with noisy signals.

There are at least two possible explanations of these contrary results. One is simply that the initial conditions (i.e. the pseudo-random number seed) may play a role in determining the quality of solution. The origin of the negative results could be some arbitrary property peculiar to the seed and network architecture. Another possible cause has to do with the standardization of the noise parameters. Standard training and validation noise levels were set to those which produced the largest effect on the Bottom, 4 hidden node, frequency domain networks. The network showing the negative results is an Air network, trained from a different pseudo-random number seed. There is no guarantee that the same parameters will cause the same effect in these two cases. Additional experiments with different training and validation noise levels and pseudo-random seeds would be necessary to determine the cause of the negative results.

Excepting the results in this one instance, the effect on network performance of training with noisy signals was to enhance the networks' abilities to classify signals with noise added to them correctly. In some cases, the networks trained with noisy signals did not perform as well on clean signals. This usually occurred in the 0 and 2 hidden node networks. In some cases the improvement persisted for testing noise levels at least as high as 1.0.

Figure 2(a): Fr[illegible]iency Domain

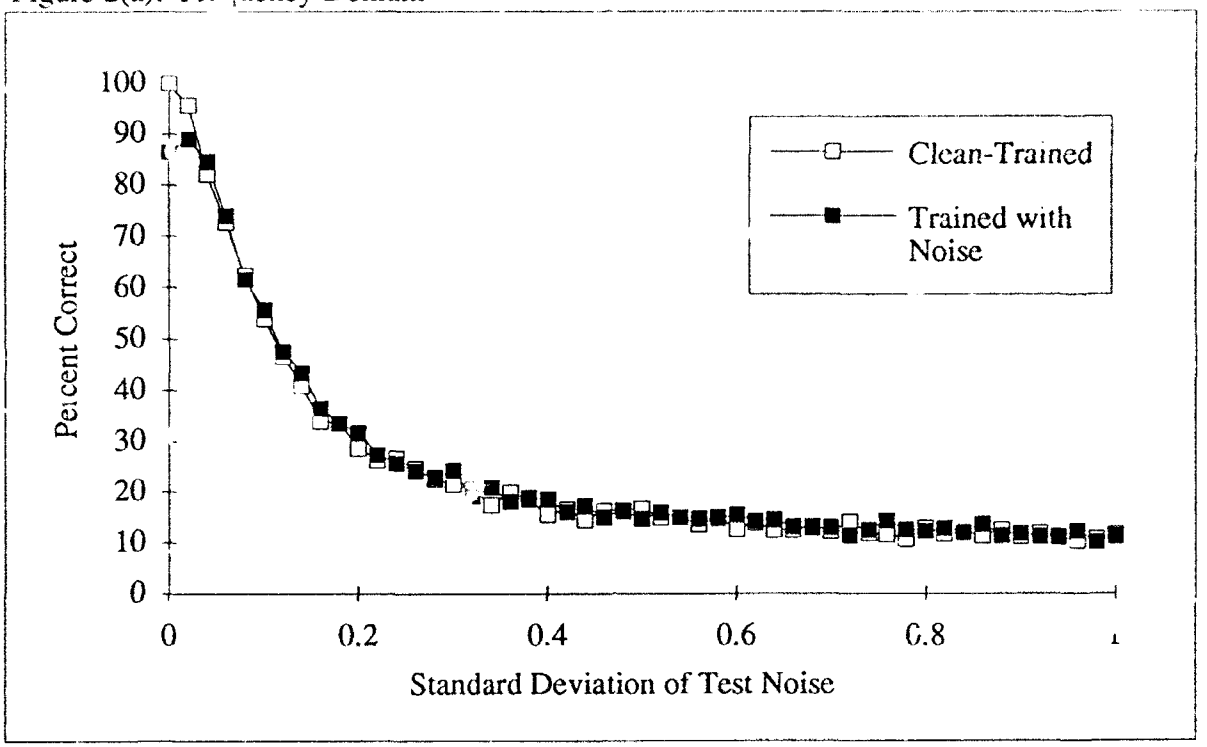


Figure 2(b): Time Domain

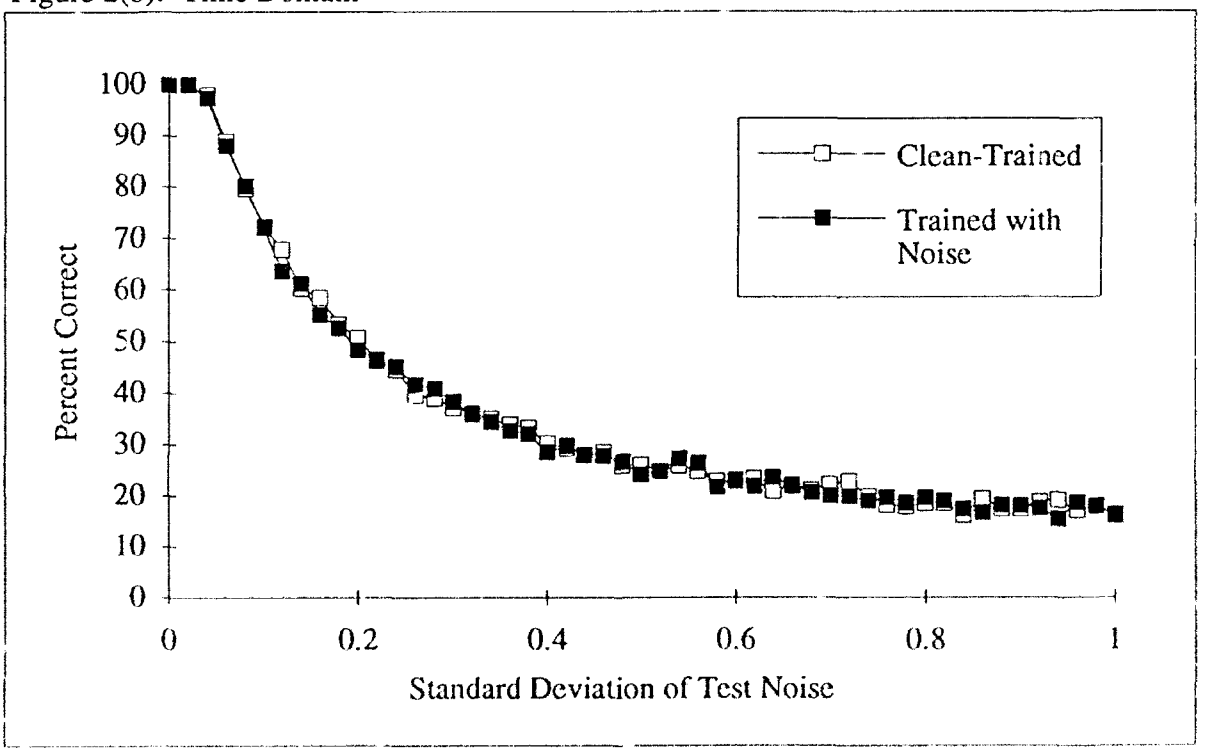


Figure 8.1.6-2 Comparison of Bot0H(2) Performance When Trained with Clean and Noisy Input Data

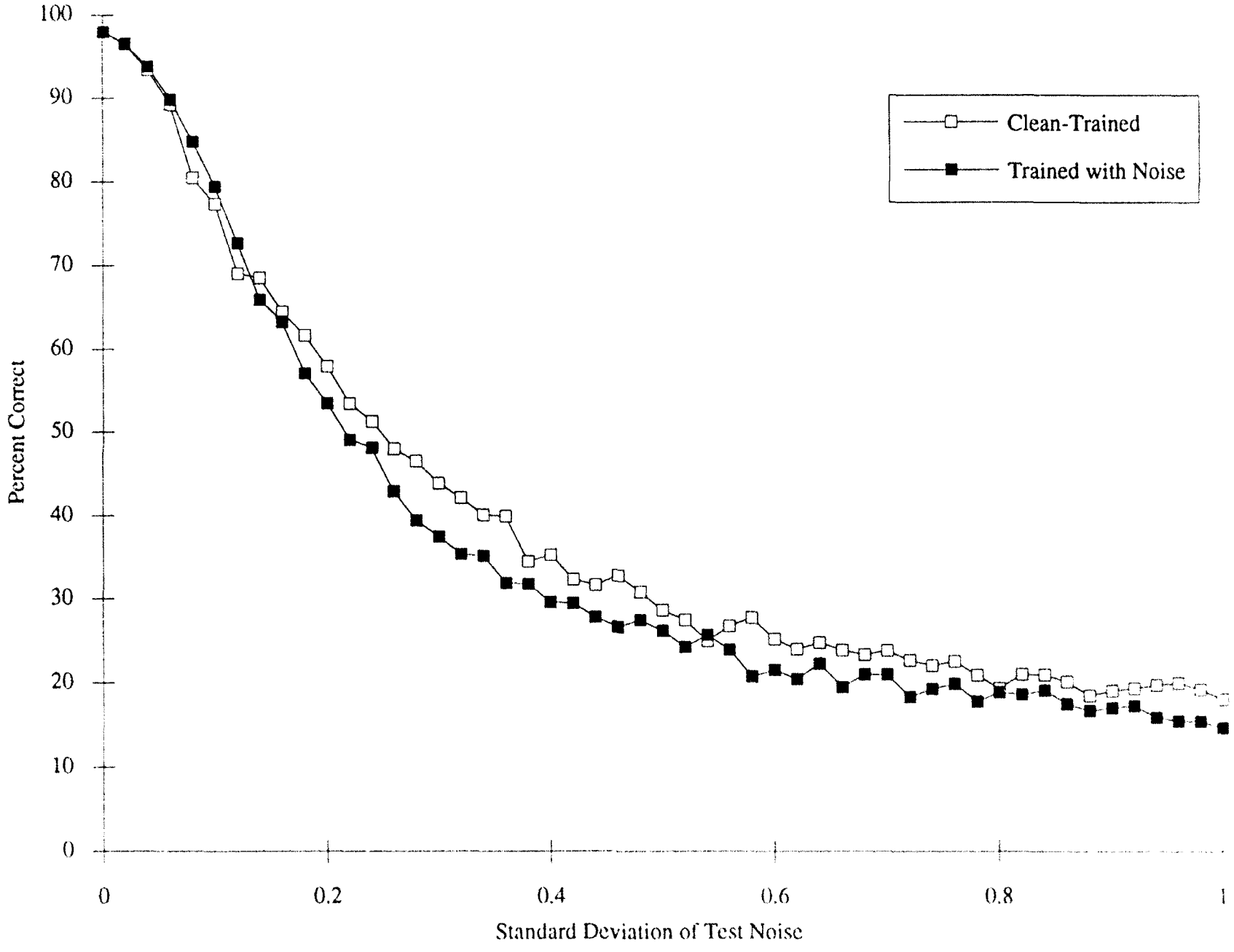


Figure 8.1.6-3 Effect on Air4H(2) Frequency Domain Performance as Noise Level on Training Data is Increased

8.1.7 Summary for Backpropagation Networks

The performance of the backpropagation networks was very high for properly configured and trained networks. Bottom and Free-field networks performed much better than subjects on the same tasks, in part due to advantages in the input representations of the signals and the networks' ability to discern detailed differences in those representations. Air networks also did well, and showed the same tendency as the subjects to have the most difficulty judging Striker. Adding artificial random noise to the signals applied to a network during training usually improved the performance of the network on noisy signals.

While networks with no hidden nodes performed well, they did so with many more parameters than other networks, allowing more arbitrary classification schemes. Networks with four hidden nodes did well on the classification tasks with relatively few parameters, and were selected for further analysis.

8.2 INTEGRATOR GATEWAY NETWORKS

Another network used to process the signal data was an integrator gateway network (IGN). Its processing is similar to that in the backpropagation network (BPN), but it is fundamentally different in the way in which it handles incoming data. The IGN has front-end layers that allow it to accumulate the values from successive patterns of incoming data and feed the accumulated data through the backpropagation-like portion of the network. The use of this type of network is driven by the need to evaluate information as it changes over time. It is particularly useful for data such as spectrograms which contain frequency information over time, and is a unique approach to network training used by Moore, Roitblat, et. al in their research on dolphin echolocation[4].

8.2.1 Network Architecture

Moore and his colleagues used the IGN on the principle that dolphins accumulate information while echolocating and identifying objects, and use the sum of what they've heard to make the identification. In much the same manner, the IGNs are used here to process spectrogram data, or frequency information in the signals over time. The structure of the network in Figure 8.2.1-1 shows an input layer, three data preprocessing layers, and a hidden and an output layer such as

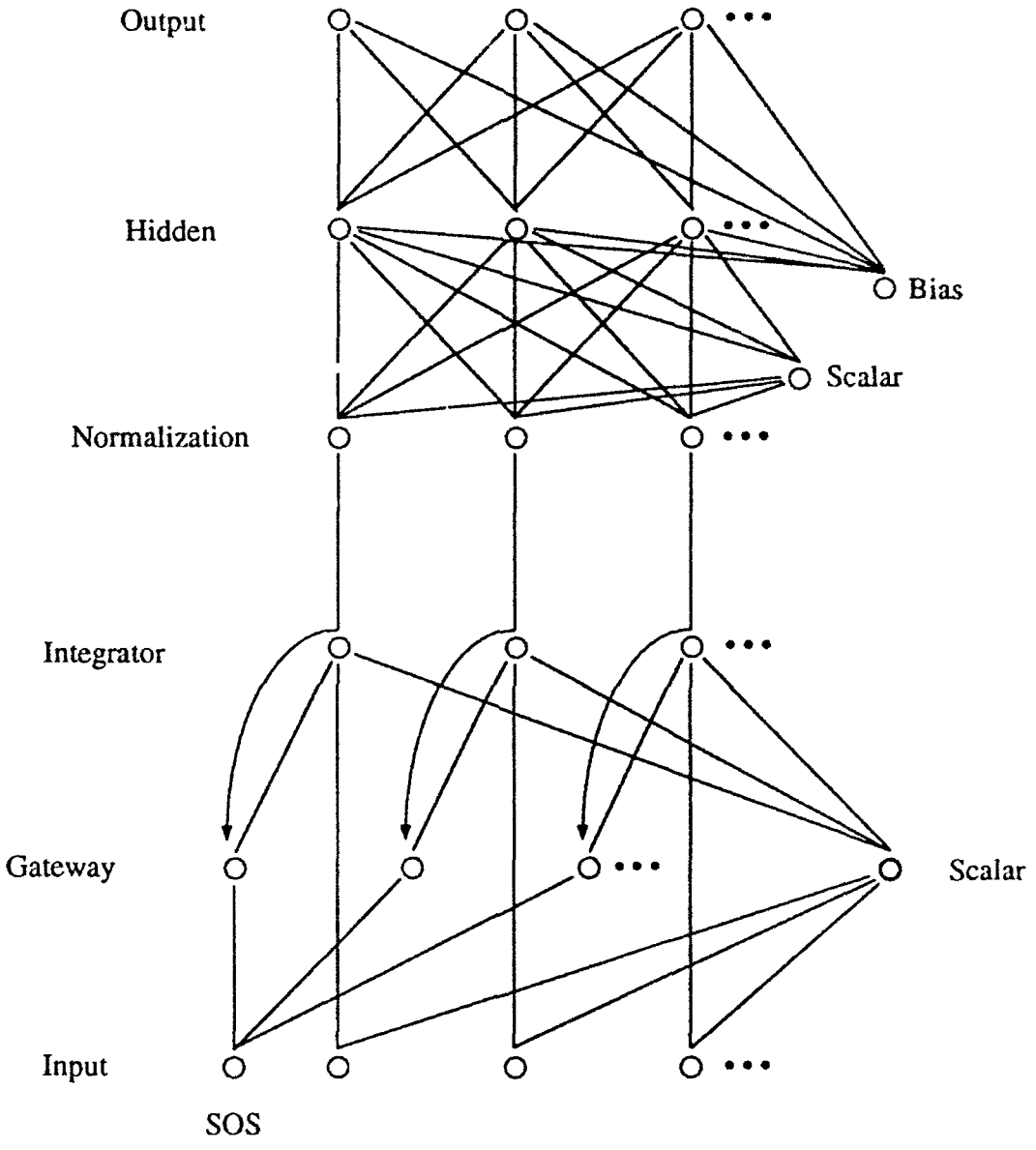


Figure 8.2.1-1 Integrator Gateway Network Architecture

those found in BPNs. Two scalar nodes are also present which are used in scaling the data from layer to layer.

For the IGNs each input pattern is considered to be a portion of a stream of patterns. The signals are presented to the networks as windows of a spectrogram created by taking a Fast Fourier Transform (FFT) as a sliding window is passed over the signal. All windows of frequency data in one signal are considered to be patterns in one stream. For this reason a data pattern is introduced to the network at the input layer and is fed to the integrator layer where it is added to the activation from previous patterns in its stream. The cumulative values then are passed through the normalizing layer where they are treated as a vector and fit to the unit circle. This normalization process controls the activation levels that will be introduced to the hidden layer where the squashing function must not become saturated. Originally the scalar node between the input and integrator layers was meant to prevent saturation, but with this particular application the activation from the signals as time progresses is too high for this scalar to handle sufficiently. The hidden and output layers function as in a general backpropagation network, with the simple addition of another scalar which controls the activation levels going into the output layer. The same target pattern is used for all patterns accumulated from a single stream of inputs. The accumulation in the network is reset at the start of each new stream by the start of stream marker. This marker affects the stream summation produced by the integrator and gateway layers.

During training the processing for the IGN was accomplished using the summation methods shown in Table 8.2.1-1. The sigmoid function was used as the squashing function on both the hidden and output layers. The accumulation processing of the IGN required that the input patterns in each stream be presented in non-random order. For this reason the cumulative delta rule was used for updating the weights on the hidden and output layers of the network. Using this rule the weights were updated each time all patterns in the training set had been presented to the network.

8.2.2 Signal Input

The accumulation nature of the integrator gateway network structure lent itself to training on spectrogram data for both the Bottom reflection and the Air signals. The spectrograms were created for both signal conditions using the time domain signals described earlier in Sections 8.1.1.2.2 and 8.1.1.2.3. For both conditions the spectrograms were generated by moving a sliding window across the signal, applying a Hamming window filter, and taking an FFT of the

resulting points. The sliding window was advanced half of the window's width for each section of the spectrogram. Some details differed for the two conditions and will now be detailed.

Layer	Summation Type
Input	Summation
Scalar	Summation
Hidden	Summation
Output	Summation
Gateway	Summation of Products
Integrator	Cumulative Summation
Normalization	Normalizing Multiplicative (fits vector to the unit circle)

Table 8.2.1-1 Summation Types Used For IGN Layers

Each of the Bottom time domain signals aligned for the back of the box was 1333 points in length. The sliding window was 64 points wide and was advanced 32 points at a time. Taking an FFT of one window resulted in 32 unique frequency values. Due to the filtering that had been performed, which was detailed in Section 8.1.1.2.2, bins 1-6 and 19-32 were excluded from the resulting data for each window. This provided 12 frequency amplitude values for use as the input for one data pattern in a Bottom signal's stream. Since all of the Bottom signals were of the same length, this method resulted in signal streams consisting of 42 time windows of frequency data.

The same type of processing was applied to the Air time domain signals with some minor changes. The duration of the Air signals of up to 32456 points dictated that the sliding window for this process be increased to 512 points. In this case the window was advanced 256 points at a time. The FFT then produced 256 frequency amplitude values. It was desirable to have an input with fewer than 50 nodes, so the 256 bin values were averaged every 8 values. This procedure resulted in 32 frequency bin values per input pattern. The variation in the duration of the Air signals resulted in streams with between 33 and 126 time windows of frequency data per signal.

The training set for the Bottom IGN contained signals which were of equal duration. Therefore, one spectrogram from each signal class for each individual instance 1-8 was included in the training set. The Air signals, however, were of greatly differing durations. In order to represent each signal class equally, the shorter signals were repeated in the training set. In other words, the longest signal had its instances 1-8 included once in the training set. The other signals' durations were compared to the longest signal's and a threshold of 65% was used to determine how many

repetitions of each shorter signal was to be included in the training set. Table 8.2.2-1 shows each signal class, the number of sliding windows in its spectrogram, and the number of repetitions of each instance included in training.

Class	Number of Windows	Repetitions for Training
B1M	41	3
B1P	33	4
B1W	33	4
B5M	126	1
B5P	125	1
B5W	108	1
S1M	121	1
S1P	95	1
S1W	84	1
S5M	67	2
S5P	67	2
S5W	64	2

Table 8.2.2-1 Input Window Repetitions For Bottom IGN

8.2.3 Network Training

The networks were trained with varying random number seeds, learning and momentum rates, and numbers of hidden nodes. The Bottom reflection processing produced signals with 12 points per window, thus the Bottom networks had 12 input nodes. In the Air data the 64 point sliding window produced 32 point FFTs which dictated that there be 32 input nodes. Each network had seven output nodes, one for each parameter value for Material, Thickness, and Angle/Striker. The output nodes for the IGNs were identical to those used in the BPNs and were shown in Table 8.1.2-3. The target values also functioned in the same way. A target of one for an output node meant that the window of spectrogram input belonged to a signal with that parameter. For example an S50 class signal had targets of one on its Steel, 5%, and 0° output nodes. The targets for the remaining four nodes then were 0.

Compared to typical backpropagation networks the integrator gateway networks with spectrogram data as input required an inordinately large number of iterations for their performance levels to peak. The differences were attributed mainly to the input data format. Many input patterns, i.e. iterations, were required to represent a single instance of a signal. Even for the Bottom set where

the signals had only 42 windows of data per signal this meant that 4032 iterations (42*12*8, windows x classes x instances) were presented to the network before a weight adjustment could be made. For the Air networks the weights were adjusted every 11536 iterations. From these numbers it is easy to see why a very large number of iterations were necessary for the network to achieve level classification performance.

These networks also required very small learning rates. A typical learning rate for a BPN was on the order of 0.1. The IGNs which performed above chance could only tolerate learning rates under 0.01, while rates under 0.005 usually proved to be most successful. It was judged that large learning rates affected the weights badly because such a large amount information was accumulated on the different windows of signals over the entire training set before the weights were adjusted. When the small learning rates were used, momentum rates more typical of BPNs were used successfully with the IGNs.

Due to the large number of iterations involved in training, the networks often required many hours to achieve above-chance classification performance for the individual parameters. This necessarily limited the number of different networks feasible to be attempted. The original approach involved running a small set of networks with 2, 4, 6, and 8 hidden nodes. In the interest of time, once it was discovered that the 6 and 8 hidden node networks did not improve the performance largely over those with 4 hidden nodes, the remainder of the networks run used 4 hidden nodes. For the Bottom condition 30 networks were trained and for the Air signals 18 networks were trained.

The performance for each of the parameters Material, Thickness, and Angle/Striker, as well as the MSE tended to fluctuate during training. In other words, it was rare that a network tested every 500,000 iterations showed a consistent increase in its percent correct for each of the parameters, as well as a steady decrease in the MSE overall. This reason, combined with the fact that the networks took a large amount of time to train, led the researchers to stop training when it was judged that the percent correct for the individual parameters had peaked or leveled.

The networks were tested against instances 9-16 of each of the 12 signal classes and their performances recorded. The tests consisted of presenting each window of each signal to a network and recording the network's response for each of the parameters. The percent correct was then computed for each parameter, as well as for the case where the three parameters had to be correct simultaneously in order for the overall measure for the signal to be correct. The network was

judged to have a correct classification of a parameter when the output node corresponding to the signal's actual parameter was the highest for all nodes corresponding to that parameter. For example, if the first window from the tenth instance of the B54 signal class were presented to the network, the response from the Brass output node would have to be higher than that from the Steel output node in order for the network to have a correct Material classification for window one of the sixth B54 signal. From this data the percentages and MSE were computed for: each window (collapsed across signal instances and classes), each signal instance (collapsed across windows), each signal class (collapsed across windows and signal instances), and the entire network (collapsed across windows, and signal instances and classes). These different measures of performance are explored in more detail in the following sections.

The most successful Bottom reflection network had 4 hidden nodes, was trained with a learning rate of 0.005 on the hidden layer and 0.003 on the output layer, and with a momentum factor of 0.3 on the hidden layer and 0.2 on the output layer. It was trained for 8,500,000 iterations where each iteration included the presentation of one input pattern. The most successful Air signal network also had 4 hidden nodes, but its learning rate was 0.003 for the hidden layer and 0.001 for the output, and it had no momentum factor on either the hidden or output layers. It was trained for 10,000,000 iterations of the Air signal patterns. The results of each of these best performing networks is discussed below.

8.2.4 Results

The unique presentation of the signals as windows from spectrograms changes the manner in which the networks' performance is evaluated. Typically a network's overall performance by parameter is the means by which it is judged. Here, the performance measures for each parameter can be viewed from an overall perspective or relative to the individual spectrogram windows. In each case, the performance computations for the Material, Thickness, and Angle or Striker parameters are collapsed across windows, and signal classes and instances in the test set, as well as for the network as a whole. Also, the mean squared error (MSE) measures the average error per output node of either a window, class, instance, or the entire network.

The cumulative processing of the IGNs lends itself to the concept that the network should perform at chance levels until enough windows from a stream have been presented that there is sufficient information accumulated in the network from which a judgment can be made. In other words, as

more windows from a stream are presented to the network, it has more information on which to base its identification of the parameters. The performance levels for the Air and Bottom networks differ greatly both by network and by where in the sequence of windows they perform well.

8.2.4.1 Air Signals

It is interesting then to note that the best performing network trained with the Air signal spectrograms achieves perfect performance on Material and Thickness at the first window's presentation, and maintains that performance across all windows. From this it can be assumed that there is information even at the beginning of each signal that captures the essence of Material and Thickness, and thus allows the network to make correct identifications with only one window's frequency information. One explanation for the network's unexpectedly fast identification involves the Air signals themselves. These signals are aligned by their initial speculars in which the energy is caused by the Striker contacting the target. The 64 point window of the signal used to produce the first 32 point spectrum input pattern thus contains a large amount of resulting signal energy. It is proposed that this impact energy contains enough information for the network to correctly identify the Material and Thickness of the signal.

Conversely, the Air network's performance on Striker is lower and less consistent. It achieves its maximum correct identification percentage of 73% for the Striker parameter by the 13th window (of 126 total windows), but does not maintain it. Thereafter, performance slowly decreases to a level of 66%. The network's MSE is at its lowest of .062 at window ten and gradually increases as the Striker performance decreases to .074. There are particular Strikers which are consistently difficult for the network to identify while others are classified correctly for 85-100% of the tests. The performance on Plastic Striker for Brass targets is a negligible 1% and 6% respectively for targets with 10% and 5% shell thicknesses. Likewise the network never (0%) identifies the Striker as Metal for Steel targets with a 5% shell. Although the performance of 61% for Metal striker on a Steel 10% shell target is above the statistically significant level of 43.75%, it still indicates that the network struggles with this classification. Overall, though, the performance for Striker is 67.5%, which is significantly above chance. The performance values on Striker for this network are similar to those from the backpropagation networks trained with both time and frequency domain, although the Strikers with which the different networks have difficulty vary. Given that their overall performance is lower and less consistent, the three Best human performers also have more trouble identifying Striker than they do Material and Thickness.

A theory about the decrease in performance by window for the Air IGN involves the idea that the majority of the frequency information from the Striker impact is available only in a set of several windows at the beginning of each signal. Although this information is retained in the accumulation of frequency energy over the life of the signal and in the way in which the target vibrates, its contribution to the overall frequency content becomes significantly lower in proportion to the target reverberation energy as the windows progress. While Striker performance does fall, the overall level is 67.5% and the decrease is gradual. Since the network performs statistically above the chance level of 33.3% it can be assumed that it retains and can identify information about the striker type throughout the set of input windows.

	Material	Thickness	Striker	Overall	MSE
Air	100.0	100.0	67.5	67.5	0.071
Chance	50.0	50.0	33.33	8.33	NA
Significant Levels	61.0	61.0	43.75	16.67	NA

Table 8.2.4.1-1 Average Air IGN Performance Compared to Chance Levels

The Air network's overall performance levels are shown in Table 8.2.4.1-1. It is of interest that Striker proves to be the most difficult parameter considering the results for the Bottom reflection networks and experiments discussed in other portions of the report. Comparing Air results to those based on Bottom data the findings show that for the underwater signals Angle is easier to distinguish than Material and Thickness. Although Striker is not parallel to Angle in the classification task, due to the radically different collection environments, the difference in performance is still notable. Remember that all of the signals were created using the same physical targets so they share the same Material and Thickness characteristics. The point here is simply that, regardless of the common targets, the networks are not able to learn Material and Thickness to the same degree for the Bottom and Air signal conditions. It is difficult to conclude whether the difference stems from Angle characteristics being innately easier to hear or from the Striker being so difficult to discern that the solutions are concentrated on the Material and Thickness distinctions.

8.2.4.2 Bottom Signals

The performance of the best network trained with Bottom reflection data is markedly different and less straight-forward than that for the Air-trained network. For this reason, it is investigated on a more detailed level. Its particular trends in performance by window are examined. Also, in order to compare the Bottom network's performance to that of the human subjects the test signals' windowed output data is scaled and the resulting dimensions are compared to those from the human scaling solutions.

The Bottom reflection data integrator gateway network (IGN) has performance levels which are significantly above chance for all parameters separately, as well as for the three parameters together which is referred to as the overall condition. The percent of correct identifications follows in Table 8.2.4.2-1.

	Material	Thickness	Angle	Overall	MSE
Bottom	67.4	64.9	76.0	37.5	0.180
Chance	50.0	50.0	33.33	8.33	NA
Significant Levels	61.0	61.0	43.75	16.67	NA

Table 8.2.4.2-1 Average Bottom IGN Performance Compared to Chance Levels

These numbers are based on the testing methods described above where the test set consists of instances 9-16 of the 12 signal classes. The signals consist of windowed spectrogram data as before and there are 42 windows in each signal. In particular this section will concentrate on examining the performance by window, and the resulting data as it is used as input to multidimensional scaling algorithms, and compared to the dimensions from human data scaling solutions.

As in the case of the Air network, the Bottom IGN is less successful on certain parameters for given signal classes than for others. The details of this are readily apparent in Table 8.2.4.2-2 which shows percent correct and MSE for parameters collapsed across windows and test instances giving performance by signal class. Note that the Material and Thickness performances on class S10 are particularly low, and that four classes have a 0% overall success rate. These low figures

imply that although the network has learned features of the signals which indicate Brass or Steel, the S10 class contains the Brass features and thus is often misclassified. The results will be discussed in further detail from the perspective of classification percent correct by window.

The performance of the network as the spectrogram windows progress shows expected as well as unexpected results. The overall trend of the performance is expected to be near chance levels until the network receives enough information in a stream to determine the parameters associated with that stream's signal class. After that, it is reasonable to expect the performance to increase as more windows' information is added to the network's accumulation for that stream. At some point, the new information available in the signal's energy will taper off relative to the overall stream's energy, thus the network's performance can be expected to level off in the later windows.

Class	Material	Thickness	Angle	Overall	MSE
B10	0.99	0.51	0.83	0.35	0.174
B14	0.89	0.90	0.87	0.77	0.101
B19	1.00	0.75	0.74	0.74	0.073
B50	0.96	0.61	0.91	0.57	0.166
B54	0.89	0.38	0.54	0.00	0.214
B59	1.00	0.26	0.74	0.00	0.178
S10	0.02	0.07	0.97	0.00	0.253
S14	0.25	0.90	0.86	0.25	0.243
S19	0.64	0.75	0.71	0.60	0.166
S50	0.50	0.97	0.97	0.50	0.188
S54	0.24	0.81	0.16	0.00	0.278
S59	0.71	0.88	0.80	0.71	0.122

Table 8.2.4.2-2 Bottom IGN Performance by Class Across Windows

Closer observation of the network's performance reveals unusual values for the Angle parameter in the first ten windows. It is important to remember that the bottom reflection data contains just that, bottom reflection, and the actual energy from the target return is not part of the signals until approximately the eleventh window of data. This can be seen most clearly in Figure 8.2.4.2-1 in the comparison of a signal containing only bottom reflection data to a B19 class signal in which the target's energy is embedded in the bottom return. Figure 8.2.4.2-2 shows that the average percent correct for Angle in the first 11 windows is 55%, while chance performance is 33.3%. Investigation of this phenomenon requires observing the performance for each of the 0°, 45°, and 90° angles. Their performance across windows can be seen in Figure 8.2.4.2-3. The interesting

Figure 1(a): B19

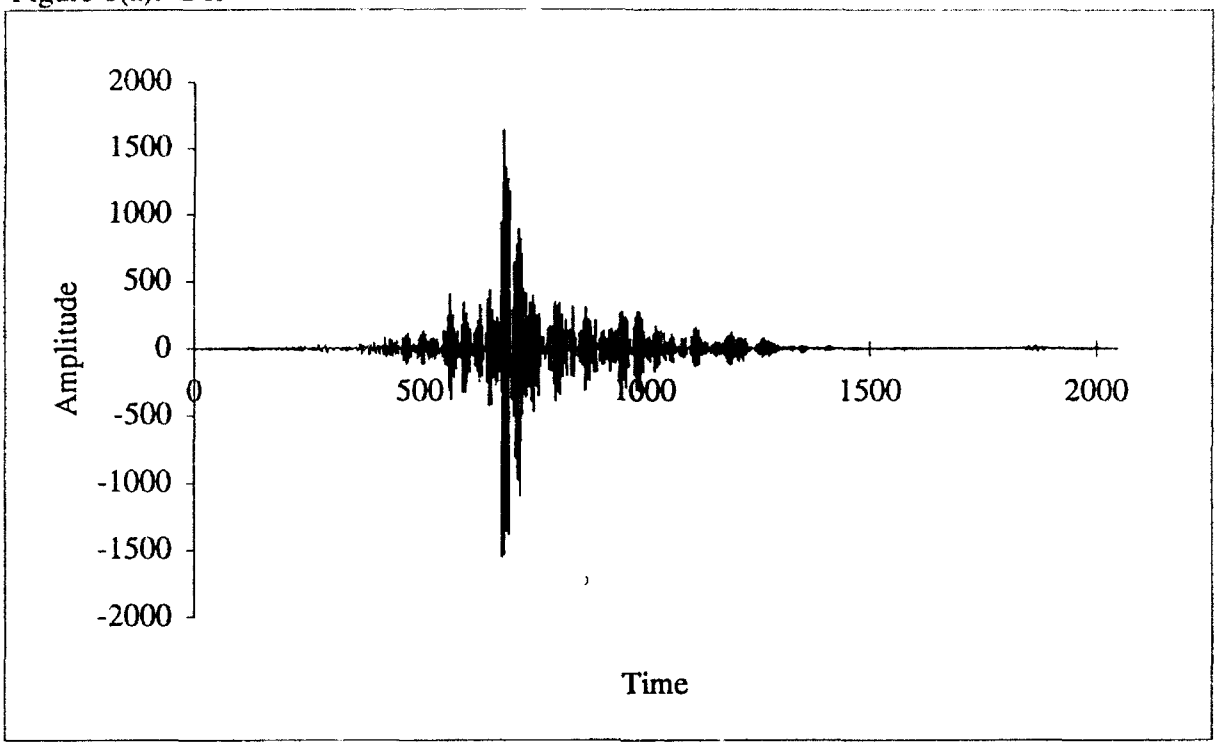


Figure 1(b): Bottom Only

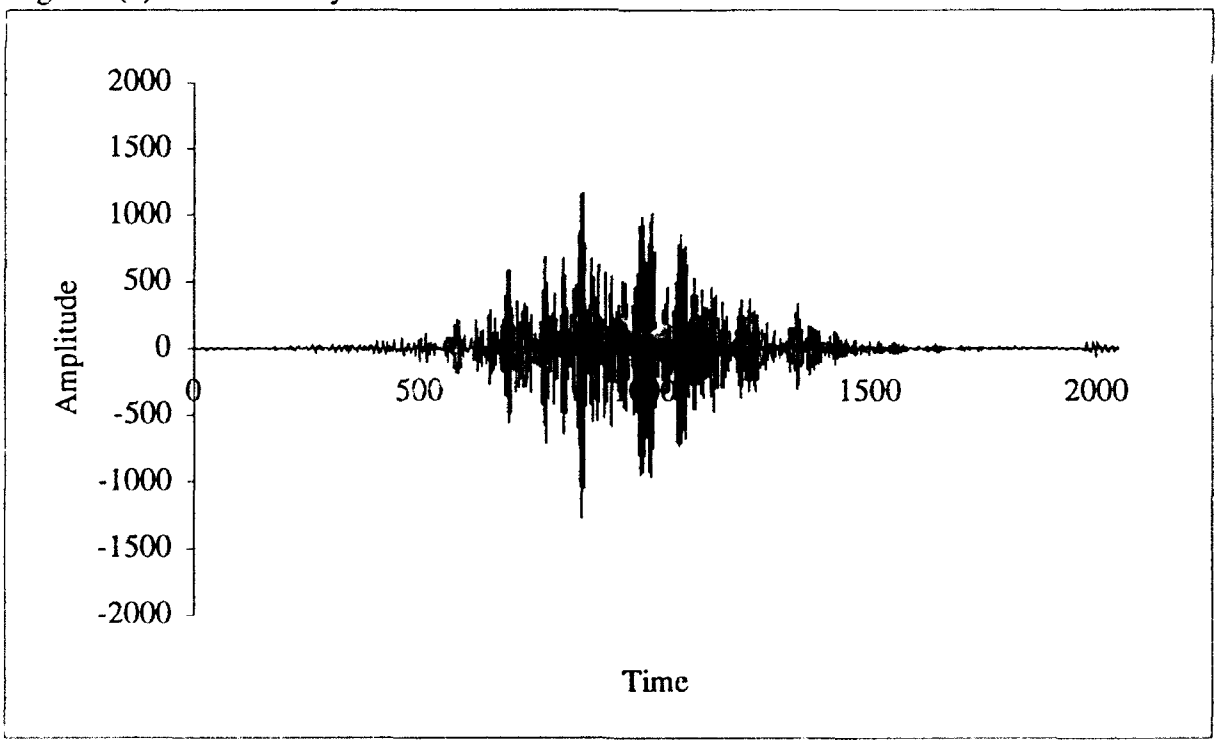


Figure 8.2.4.2-1 Averaged Signals B19 and Bottom Reflection Only

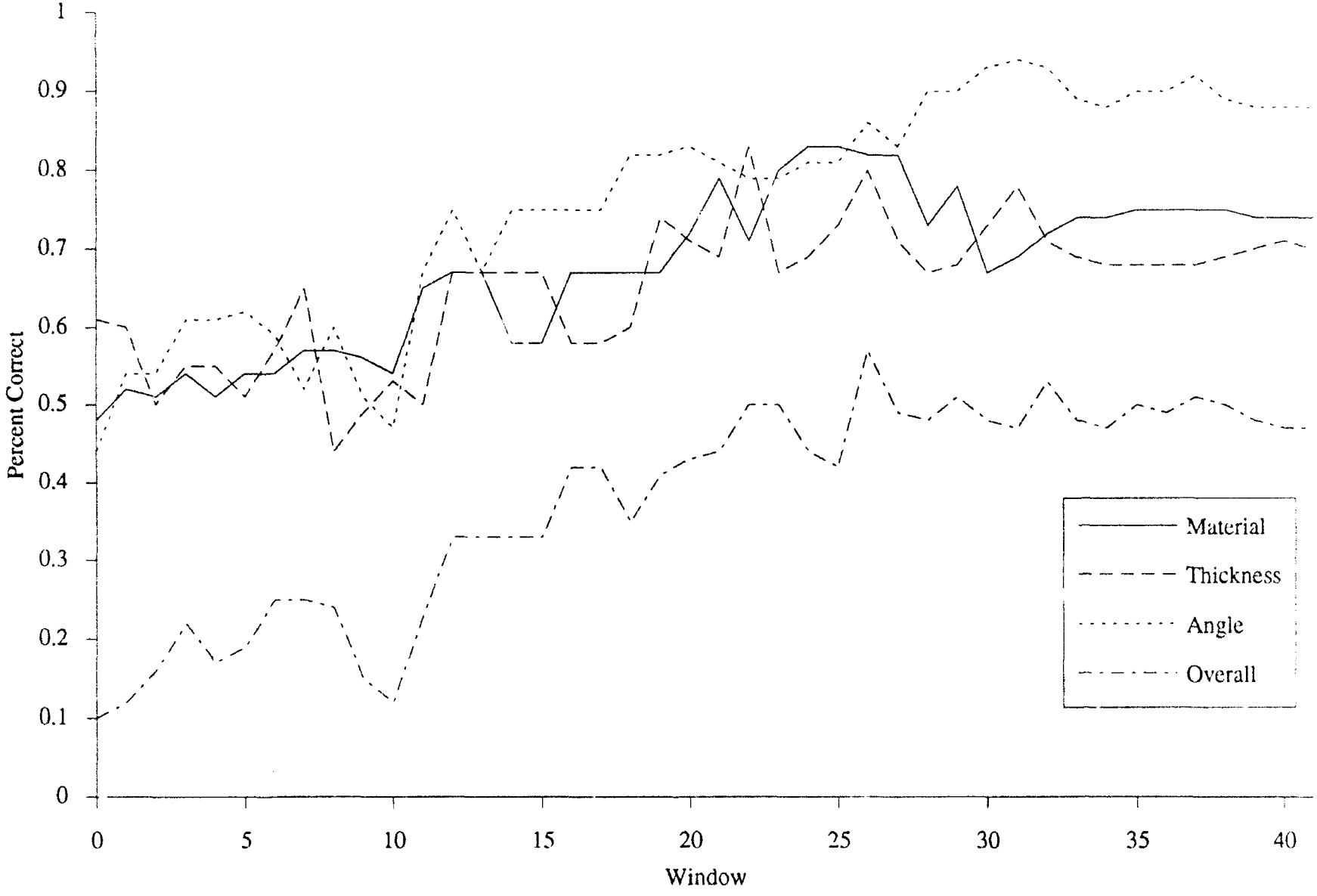


Figure 8.2.4.2-2 Bottom IGN Performance by Window

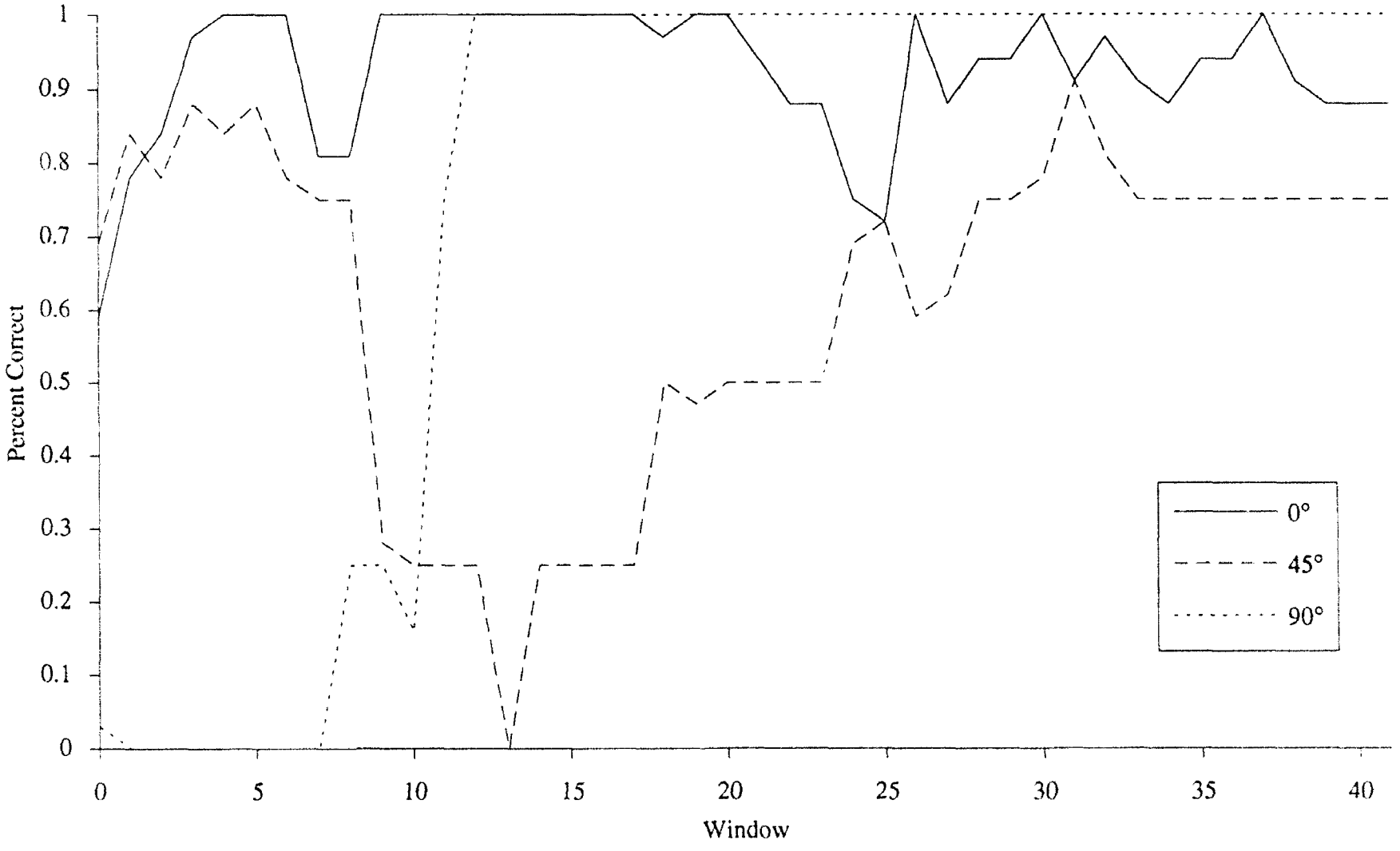


Figure 8.2.4.2-3 Bottom IGN Angle Performance by Window

aberration in this graph is that the performance for both the 0° and 45° signals is well above chance until window eleven, although the target return is not present in the signal at that point. After that the 45° performance drops dramatically, while the 0° signals take a small, but relatively insignificant dip. The 90° signals' percentages do not follow the expected chance performance trend in their first eleven windows either. The network is classifying almost all of the initial 90° signal windows as being from 0° signals, instead of randomly "guessing" their true identity. In some way the network has learned anomalies about the bottom reflection portion of the signals that allow it to classify the 0° and 45°, but not the 90°, signals. For this reason the performance is above the expected level of chance in the first several windows.

Once the network gets beyond the first windows, it begins to perform more as expected. Figure 8.2.4.2-2 shows that the performance rises for all parameters in a steady manner, and peaks by window 31 where the amount of signal energy added to the sum for a stream in the network starts to become proportionally small. This display of expected behavior makes windows 13-31, over which the performance is on average steadily increasing, a logical subset to use in comparing the Bottom network's performance to that of the human subjects. Due to the windowing nature of the spectrogram data, and therefore the results, a method of direct performance comparison is generated. It is decided that multidimensional scaling of the confusions produced by the networks over the windows of interest will be the best way of equating the results with those from the human experiments.

Scaling the results from the Bottom network involves creating confusion matrices from its resulting data. This is accomplished in the same manner as for the human subjects, and the process is described in Section 7. Each output from the network is tallied in a matrix of actual versus classified signals. In other words, if a network is given an instance of a B19 signal and identifies it as a B59 signal, the B19 row, B59 column has one added to it. After the output for all of the signals has been tallied, the matrix contains similarity data which represent the ways in which the signals are confused by the network. A confusion matrix is created for each window in the set of increasing windows 13-31. Scaling solutions are generated for several sets of windows, and their resulting dimensions are examined. The solutions are produced by running an individual differences scaling algorithm using the windows' confusion matrices as input. Since one network produced all of the confusions, the scaling is run in the "unconditional" condition. This means that the raw confusion numbers can be treated as equal from matrix to matrix. The solutions produced by the scaling runs are examined below.

The three-dimensional solutions are chosen as the best comparison dimensions due to the fact that the human dimension solutions evaluated contain three dimensions. A subset of windows 13-31, including 13, 17, 21, 23, 25, and 29, are examined first. The subset's scaling dimensions are shown in Figure 8.2.4.2-4, where it can be seen that their solution is very similar on dimensions one and two to the two scaling solutions for the Best and single top performers. The dimensions from the solutions for the human performers were shown in Figures 7.4.2-1 and 7.4.2-2.

Note that the first dimension in each of the three figures is divided by the 90° signals versus the 45°/0° signals. This implies that more network and human classifiers could discern the 90° signals from all others better than they could with any other characteristic in their identification schemes. It is unimportant that the order of the signals along the dimensions appears inverted from low to high. What is important is that the relative order of the signals on the three first dimensions is similar. The Bottom IGN solution orders the 90° signals on this dimension very similarly to 90° signals on the Best Bottom first dimension. In particular, note that in both cases the S59 signal class is separated from the other three 90° classes. These two solutions also have three of four 0° signals lower on the dimensions than the 45° signals. The outlying S50 class is also closest to the middle than any of the other 45°/0° classes for both solutions.

The second dimensions for all three solutions split the signals into three separate Angle categories. As in the case of the first dimensions, the parallels among the second dimension distributions is marked. The 45° classes are at the lower end of the dimensions, the 90° classes are clustered in the middle, and the 0° signals are at the high end. Although, for the network, the B10 class was with 90° signals and S54 was with the 0° group, the similarities are still striking.

From the parallels seen in the first two dimensions for the three Bottom scaling solutions, it can be concluded that the network and the human subjects concentrate on similar features of the signals when performing the classification task. The fact that the data from both the humans and the network produced two of three dimensions devoted to Angle attests to this parameter's importance in all three solutions. The performance for the three also shows that Angle was the easiest of the parameters to identify.

Although the previously described subset's solution best matches those of the humans, the entire increasing portion of the network needs to be included in the examination in the interest of

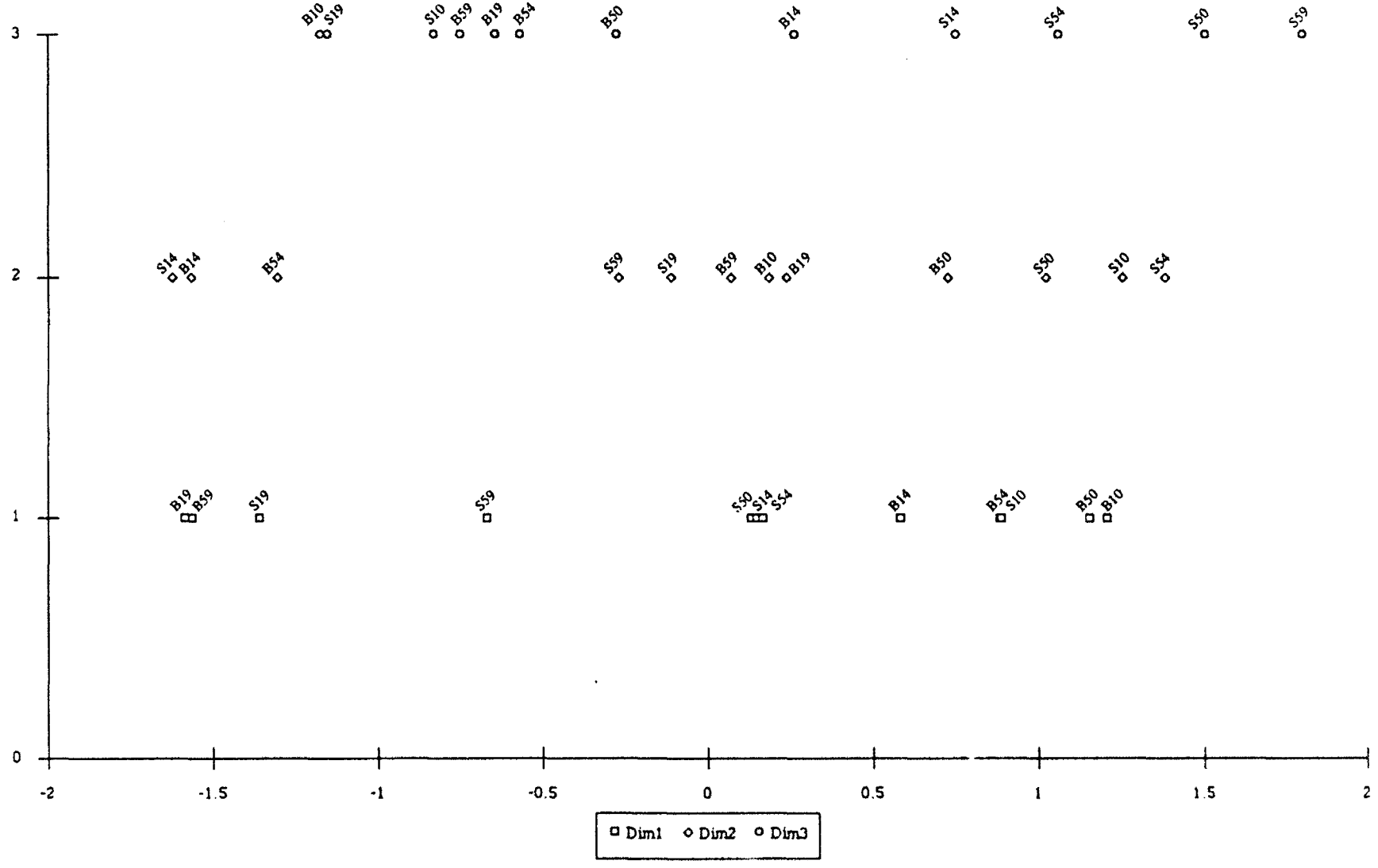


Figure 8.2.2-4 Bottom IGN Windows 13, 17, 21, 23, 25, and 29 Scaling Dimensions

thoroughness. Comparison of the dimensions for network windows 13-31, shown in Figure 8.2.4.2-5, and for the two human solutions reveals an interesting difference in their approaches. The first network dimension matches the two second human dimensions. However, one can look at the overall solutions as being more similar than would seem at first glance. Although the three first dimensions show similar signal class distributions, this is particularly true for the network and the Best performers' solution. For these two solutions S59 is separate from, although still clustered with, the other 90° signals. Also, the 45° signals are closest to the 90° signals and have S50 included with them. The other three 0° signals are at the high end of both dimensions as well. The network's second dimension is not as well separated by Angle as the two human second dimensions. For the network dimension, the 90° and 0° signals were intermixed while the human dimensions distinguished them perfectly. Even so, with the exception of the signal class S54, the placement of the 45° signals at the extreme low end of the second dimension is common to all classifiers.

Although the network scaling solution using data from windows 13-31 has remarkable similarities to the human solutions, there are also noteworthy differences. For instance the clear separation of the three Angles on the network's first dimension, which only occurs on the second human dimensions, shows that the network's output data reflects this distinction more. Also, the network's third dimension divides by Material, with the exception of the classes S19 and S10 being located among the Brass signals, while none of the human dimensions breaks down by Material. Additionally, the Steel 5% signals are at the high end of the third network dimension. The network's performance for Material and Thickness actually reflect even more of an ability to discriminate these parameters than is reflected in the separation of these parameters on the three network dimensions. In general the network and humans show common uses of signal characteristics as reflected by their scaling solutions for Angle, but not for Material and Thickness.

The investigation of the similarities between the network and human approach to the classification is continued by looking at correlations in their data. Correlation measures were computed using the values of the signals as they were distributed along the dimensions for the network and two human scaling solutions. The correlations can be seen in Table 8.2.4.2-3. The similarities seen between the first dimensions from the network's and the Best Bottom's solutions are reflected in a very high inverse correlation of -0.95. Likewise the network's and N6's first dimensions have a high inverse correlation. The relations are inverse in both cases due to the opposite ordering of the signals along the dimensions. Although 0.70 is the cutoff for statistical significance at the one

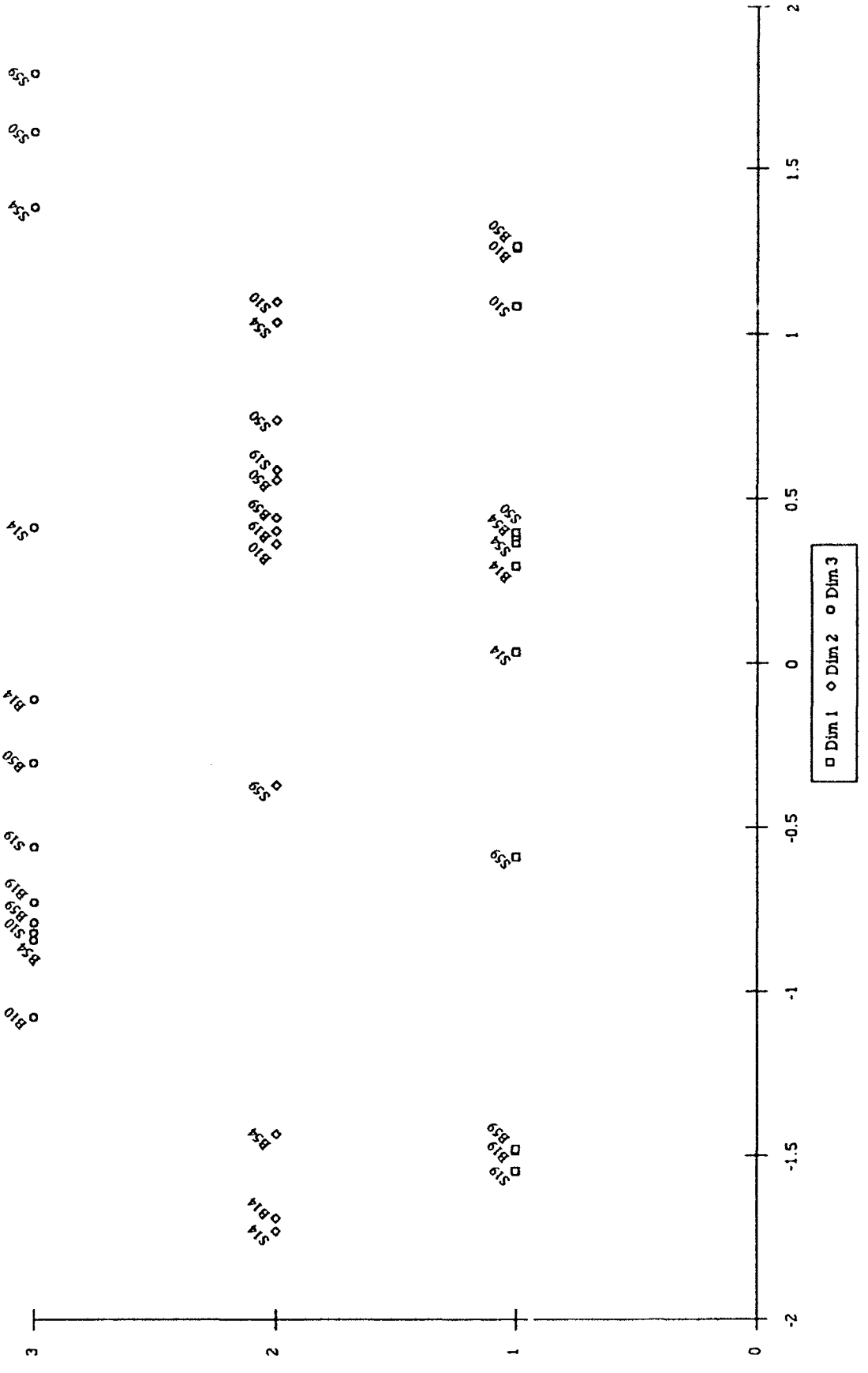


Figure 8.2.4.2-5 Bottom IGN Windows 13-31 Scaling Dimensions

percent level, the levels for the second dimensions in both cases are in the 0.6+ range. This indicates that although they aren't correlated beyond a doubt, there is a high measure of relation between them. The high correlation values for the different dimensions serve to reinforce the conclusions from the observations discussed above.

	BEST		
	BestBott Dim1	BestBott Dim2	BestBott Dim3
Win13-31 Dim1	-0.95	0.28	-0.17
Win13-31 Dim2	0.14	0.65	0.40
Win13-31 Dim3	-0.08	-0.20	-0.06

	N6		
	Bott N6 Dim1	Bott N6 Dim2	Bott N6 Dim3
Win13-31 Dim1	-0.88	0.28	0.33
Win13-31 Dim2	0.21	0.63	0.37
Win13-31 Dim3	-0.09	-0.2	-0.42

Table 8.2.4.2-3 Correlations of Bottom IGN Windows 13-31 and the Bottom Best and Subject N6's Scaling Solutions

As described in Section 7, the subject weights from the individual differences scaling solution are another way of viewing the relations between subject sets. For the Bottom integrator gateway network, the subject weights show relatively little variation in the use of the three dimensions. This is different than what is experienced in the human dimension solutions discussed earlier. The human subjects tend to use the dimensions differently, both with respect to other dimensions in their solutions and to other subjects. The network shows a consistency of dimension use that holds across "subjects," windows in this case, as well as among dimensions for one window. Table 8.2.4.2-4 does show some dimension use difference in that the angles for the first dimension are smaller, thus it is being used to a slightly greater extent than dimensions two and three. The overall importance measures for the three dimensions also vary less than those for the human solutions. It is interesting that, in general, the networks use the dimensions in a more consistent manner than the humans, yet their results are strikingly similar.

BEST

	Subject Weights			Angles			
Window	Dim1	Dim2	Dim3	Dim1	Dim2	Dim3	Weirdness
13	0.325	0.289	0.274	50.746	55.797	57.822	0.014
14	0.320	0.275	0.277	50.639	56.994	56.723	0.035
15	0.325	0.289	0.273	50.737	55.794	57.834	0.013
16	0.321	0.291	0.268	50.961	55.196	58.210	0.010
17	0.319	0.290	0.271	51.262	55.237	57.841	0.003
18	0.310	0.297	0.270	52.364	54.169	57.770	0.019
19	0.320	0.306	0.270	51.940	53.866	58.550	0.024
20	0.326	0.303	0.269	51.180	54.380	58.834	0.021
21	0.325	0.302	0.272	51.335	54.576	58.455	0.014
22	0.325	0.290	0.285	51.334	56.145	56.832	0.022
23	0.324	0.295	0.275	51.198	55.247	57.900	0.004
24	0.326	0.304	0.276	51.572	54.568	58.205	0.011
25	0.325	0.294	0.286	51.642	55.749	56.907	0.017
26	0.326	0.298	0.284	51.644	55.445	57.216	0.011
27	0.325	0.302	0.286	51.986	55.158	57.145	0.013
28	0.325	0.292	0.288	51.602	56.104	56.590	0.025
29	0.326	0.304	0.271	51.336	54.338	58.707	0.020
30	0.326	0.302	0.281	51.678	54.970	57.670	0.005
31	0.326	0.311	0.276	51.894	53.908	58.555	0.024

Overall Importance	Dim1	Dim2	Dim3
	0.105	0.088	0.077

Table 8.2.4.2-4 Bottom IGN Scaling Solution's Usage Measures

8.2.5 Summary

The IGNs examined in this section proved to be capable discriminators of parameters for both the Air and Bottom signal sets. The Air network's perfect performance on Material and Thickness is outstanding, and matches the best BPN's performance. Its 67.5% correct identification of Striker is significantly above chance, although it does not match the performance from backpropagation networks. Note that the human subjects, as well as the backpropagation networks, had the same relative success with Material, Thickness, and Striker as did the IGN. The Bottom signals showed good results as input to this type of network. The IGN performed statistically above chance for each parameter individually, as well as overall, although it did not match the perfect performance of the Bottom BPNs. One of the most interesting aspects of the IGN's performance involved its relationship with the human scaling solutions based on Bottom data. The scaling dimensions and

the Bottom IGN share several characteristics in the ways in which they approach their solutions. They each stress the Angle parameter in very similar manners. This is particularly interesting considering that time windows of frequency data were the input to the network. It gives credence to the theory that the humans are using both time and frequency domain information in performing the classification task, and shows that their approach can be mimicked by the integrator gateway networks.

9.0 SIGNAL STATISTICS

Models of the signal scaling dimensions were required for comparison to the strategies of nodes from the neural networks described in Section 8. The primary building blocks of these models were certain parameters of the signals which fell into three classes. The first class was a group of parameters computed as statistics of the frequency distribution of a signal:

Mean
Mode
Standard Deviation
Skewness
Kurtosis
Low Frequency Slope
High Frequency Slope

The second class of parameter was a pair of measures computed in the time domain:

Decay Amplitude
Decay Damping

Finally, the Air signals were also characterized by fitting a set of sine waves to the signals and taking the following parameters of those sine waves:

Curve Fit Amplitude
Curve Fit Decay Coefficient
Curve Fit Frequency
Curve Fit Phase

9.1 FREQUENCY DISTRIBUTION AND TIME DOMAIN MEASURES

The basis for the signal statistics was the frequency distribution of the signals. This was computed for each signal by first taking the Fast Fourier Transform (FFT) after a Hamming window was applied. At each resulting frequency point the real and imaginary parts were squared and the squares were summed.

$$P(i) = X_{re}(i)^2 + X_{im}(i)^2$$

The frequency distribution of a particular signal was treated as a probability density function (pdf) by dividing each point by the sum of energy at all points.

$$p(i) = P(i) / \sum_{i=1}^{n} P(i)$$

where n differs by signal category (Free-field, Bottom, Air).

The spectral moments were then computed from the pdf as follows:

$$M1 = \sum_{i=1}^{n} f(i)\, p(i)$$

$$M2 = \sum_{i=1}^{n} (f(i) - M1)^2\, p(i)$$

$$M3 = \sum_{i=1}^{n} (f(i) - M1)^3\, p(i)$$

$$M4 = \sum_{i=1}^{n} (f(i) - M1)^4\, p(i)$$

where f(i) is the frequency at point i.

The mode of the distribution is the frequency with the maximum energy. The first moment (M1) is the mean of the sample distribution, which in this case is the mean frequency. Skewness and Kurtosis are computed as:

$$\text{Skewness} = M_3 / (M_2)^{3/2}$$

$$\text{Kurtosis} = (M_4 / M_2^2) - 3.$$

The high and low frequency slopes of the distribution were computed as a means of measuring how quickly the distribution fell off from the peak frequency. Taking the energy at each bin in the range from 0 to the mode, the slope of the best-fit line was estimated by a least-squares linear

regression. This is the low frequency slope. The high frequency slope is computed in the same manner using the energies at frequencies from the mode up to the Nyquist frequency. These measures are most useful for characterizing the underwater sounds, for which the insonifying frequency of 400 kHz can be expected to be extremely close to the modal frequency of the reflected signal.

Two further measures used to characterize each signal were computed in the time domain. These measured the damping characteristics of the Free-field and Air signals. To compute the measures a signal was rectified, and the resulting positive values were low-pass filtered in the frequency domain. The filter was applied by taking the FFT of the signal, setting the magnitude of the frequencies we wished to eliminate to zero, and taking the inverse FFT. This process is described in Section 4.

The peak of the Free-field and Air signals is at the start of the signals. Starting at the peak a decaying exponential was fit to a fixed number of points in the signals by minimizing the mean squared error of the curve. This curve is characterized by its initial decay amplitude and its decay damping constant.

9.2 CURVE FIT MEASURES

Another method for extracting features from a complicated time domain signal was to fit a parametric function to the signal using standard minimization techniques to determine the values of the parameters. It was hoped that the "best" parameters so determined would correlate well with hidden node behavior and human subject results, and so afford insight into how both humans and networks classified the signals. Due to the large amount of effort required for operations of this type, the curve fitting procedure was restricted to the Air signal set. The reason for choosing these signals over the Free-field and Bottom was that the human subject dimensions for the Air signals were more complex than for Bottom or Free-field signals. A meaningful result from curve fitting to the Air signals would aid in the modeling of these dimensions more than a similar result from Free-field or Bottom signals.

Other considerations also favored the choice of the Air signal set. The Air signals showed the largest variation between different instances of the same signal class. This made feature extraction "by eye" more difficult, and the algorithms developed by networks more subtle. Curve fit

parameters could be used to help clarify qualitative differences and similarities between the signal classes. The Air signals were also the longest signals of the three sets, with the largest signal-to-noise ratio, and thus contained the most detailed information. A carefully chosen fitting function could condense and extract such information, capturing details which were averaged away by other analytical procedures. A good result from a curve fit could be used to generate a fairly accurate approximation to the original signal. In this sense it was a "more accurate" means of extracting information.

At the outset, it seemed that finding a form for the fitting function would be difficult in the case of the Air signals, due to their long length. In general, the longer a data series, the larger the number of parameters needed to fit the data well. The introduction of more parameters ultimately would cause problems with the convergence, stability, and interpretation of the fit results, however. It was also desirable to find a form for the fitting function in which the parameters had some intrinsic physical meaning.

Fortunately, two qualities of the Air signals simplified the choice of form. First, the Air signals all began with the impact of the striker on the target, and ended when the resulting ring decayed away. This suggested the use of a fitting function with an exponentially decaying envelope. Second, Fast Fourier Transforms (FFTs) of the Air signals revealed that all of them had a significant portion of their energy concentrated in one to three relatively sharp peaks. This suggested that a fair approximation to the signal might result from a sum of a few damped sinusoids.

In addition to these purely pragmatic motivations, this choice of form for the fitting function had an appealing physical interpretation. The target, like all physical objects, had a natural set of modes of vibration, each of which had its own decay characteristics. Depending on characteristics of the striker's impact with the target, these modes of vibration were excited to a greater or lesser extent, then decayed in time. Although the number of modes was infinite, the number of modes to be excited significantly by the striker may have been small. The process of finding the best fit could therefore be thought of as a means of determining and characterizing the most significant modes of vibration excited in each signal.

The exact mathematical form of the fitting function chosen was a sum of n damped sinusoids, each of which was characterized by four real-valued parameters: an amplitude A_j, decay coefficient β_j, frequency ν_j, and phase ϕ_j ($1 \leq j \leq n$). Fits were tried using between two and six damped sinusoid

terms ($2 \leq n \leq 6$), with mixed results. The best approximation yielded by two damped sinusoids was very poor. As the number of damped sinusoids was increased, up to five, the quality of the best approximations improved. With six terms, the quality of approximations did not improve over that obtained with five, and the incidence of singular matrices became noticeably higher. Moreover, with six damped sinusoids, there was greater variation in the best coefficients for fits to different instances within a given signal class. This suggested that six terms allowed the fit to "wander" too much in parameter space, finding solutions which were not physically relevant. It was therefore decided that five damped sinusoids was the optimal number to use, with the possible exception of using a still larger number than six. The fitting function which was finally used was therefore given by the expression:

$$f(t) = \sum A_j e^{-\beta_j t} \cos(2\pi \nu_j t + \phi_j) \quad \text{for } j = 1 \text{ to } 5$$

This expression contains twenty independent parameters, whose values had to be simultaneously determined by the fitting procedure.

A standard procedure, the Levenberg-Marquardt method, was used to determine values of the parameters yielding the best fit. This procedure iteratively found values of the parameters which minimized the fit's chi-square value. The procedure was implemented in the C programming language, based very closely on published routines[5]. With the basic technique and fitting function specified, two issues remained to be addressed. First, the path in parameter space taken by any fitting procedure was sensitive to initial values of the parameters. To have confidence in the meaning of the "best" values determined by the procedure, a valid means of determining the first guess had to be developed. Second, the iterative procedure used to find a solution could, in principle, be continued indefinitely. It was therefore necessary to establish standard criteria for terminating the fit.

Initial guesses for the twenty fit parameters were determined from information contained in the complex-valued FFT of each signal. As stated above, the curve fit parameters consisted of five sets of four quantities: amplitude, decay coefficient, frequency and phase. When expressed in complex polar coordinates, a Fourier transform gives explicitly the amplitude and phase of spectral components, as a function of frequency. The amplitudes, frequencies and phases in the curve fit parameters could be computed from the amplitudes, frequencies and phases of selected components present in the Fourier transform of the signal. Extracting guesses for decay

coefficients from a Fourier transform was less straightforward. Fortunately, it proved adequate to set the decay coefficients to a qualitatively reasonable, but arbitrary value.

It remained then to find a means of selecting which Fourier components to use for the guesses. The basic approach was to choose five components which adequately represented the largest features present in the FFT. Many variations on this theme were tried, with their successes being rated by how closely the final fitted function approximated the signals. The most successful method selected the components from the FFT in the following way. The 16384 independent components of the FFT (the DC offset was not included) were divided into 16 contiguous blocks of 1024 frequency bins each. Within each block, the frequency component with the largest amplitude was selected. The 16 components so chosen were then placed in order of descending amplitude. The first (largest amplitude) component was used to compute the first damped sinusoid's initial values. Each subsequent , progressively smaller, component was examined in turn, and used to generate initial guesses provided that its frequency bin was not within 512 bins of the frequency bins of any of the other components already used for initial guesses. This provided a computationally efficient way of choosing 5 components which equally represented the most significant features throughout the entire spectrum.

From the 5 FFT components selected, the initial guesses were then computed as follows. The curve fit amplitudes, A_j, $(1 \leq j \leq 5)$, first were set equal to the amplitudes of the chosen FFT components, then all divided by the largest amplitude among them. Thus, the largest component was given an amplitude of 1.00, and the other amplitudes were scaled proportionally to maintain the same relationship between them. The choice to make the largest amplitude 1.00 was so the largest sinusoidal term had the same maximum value as the normalized signal itself. The decay coefficients, β_j, were all set to the same initial value, 16.0 s^{-1}. That is, each mode was initially set to decay to 1/e times its initial value in 0.0625 seconds which was within the first 1000 signal points. This value was empirically found to give stable and consistent results. The phases, ϕ_j, were set equal to the phases of the chosen FFT components, and the frequencies, ν_j, were set equal to the lowest frequency covered by the chosen frequency bin. From initial guesses produced in this way, the best fit parameters obtained approximated the signals to a high degree of accuracy.

Convergence criteria are a set of mathematical conditions which are evaluated after each iteration to determine whether to continue the iterative process, or stop and take the latest values of the parameters as the final result. Normally, the fit is considered good enough to stop the fitting

process when the chi-square parameter reaches a sufficiently low value, usually of order 1.0 per degree of freedom. However, this condition is valid only if the error in each data value (in this case, the value at each point in the signal) is accurately known. In the case of the Air signals, estimates for the errors were unknown, requiring that another means of quantifying the goodness of fit be used.

For every given signal and set of parameters, the goodness of fit was evaluated as follows. First the parameters were used to generate the fitting function, point by point, producing the approximation to the signal yielded by the fit. The residual signal was then computed by subtracting this approximation from the actual signal. The residual signal showed, point by point, the deviation of the curve fit resulting from the actual signal it modeled. By taking the ratio of the amount of energy contained in the residual signal to the amount of energy contained in the actual signal, a quantitative measure was obtained of how much signal energy was not well modeled by the fitting function. This ratio was named the lost fraction, and formed the basis for comparing the quality of different fits.

For some of the signals, it was found that the lost fraction (which was closely related to the chi-square) continued to drop, indicating that better choices for the parameters continued to be found, even after as many as 120 iterations. The drops in the lost fraction typically became very small after approximately 40 iterations, however. Since the iterative process was very slow, due to the long signal sizes, it was desirable to set an absolute limit on the number of iterations. An upper limit of 80 iterations was ultimately set; this was computationally reasonable, but sufficiently high to instill confidence that the parameters developed by the fit were meaningful.

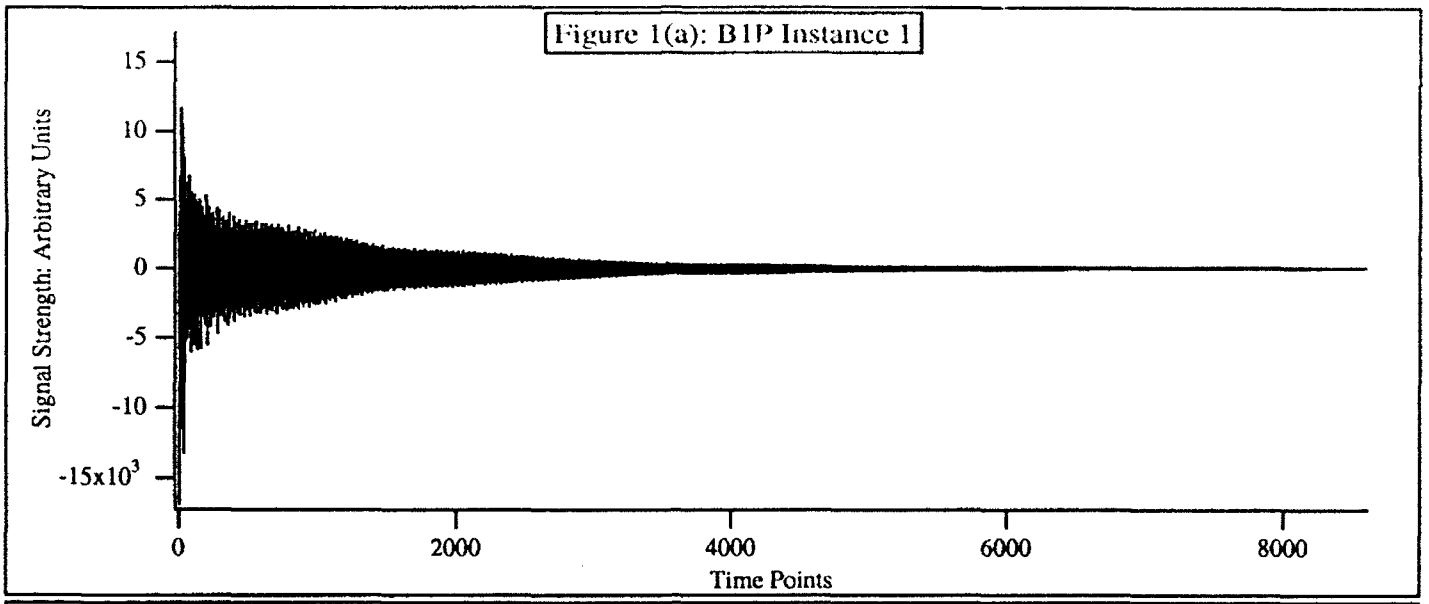
The Levenberg-Marquardt method decreased (increased) the size of the "step" in parameter space, depending on whether the chi-square decreased (remained the same) in the previous iteration. Because of this fact, it was useful to stop a fit prior to 80 iterations in the case of steps becoming either too large or too small. If the step size increased past a certain point, the changes in the parameters became too large, allowing the fit to explore parameters too far from the initial guesses to be physically relevant. To prevent this, the fit was halted if 10 iterations were completed without a drop in the chi-square. On the other hand, if the step size became too small, the quality of the results did not become suspect, but the parameters ceased to change by significant amounts, thus wasting computation time. Thus, when the step size dropped too low, the fit was stopped if

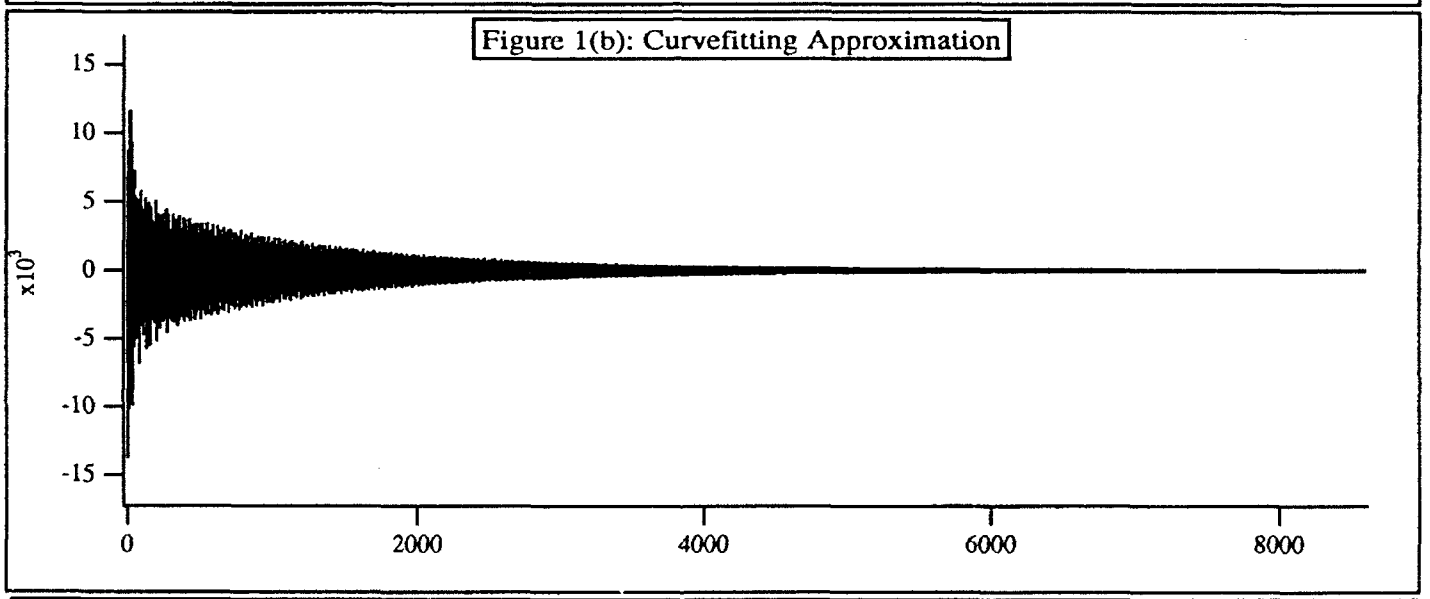
the lost fraction were less than a convergence threshold of 4%; otherwise it was reset to a moderate value, and the fit was continued.

After the fit for each signal was completed, the lost fraction typically reached a level of about 9%, with a maximum value of about 29% (instance 9 of S1W), and a minimum of about 1.7% (instance 9 of B1P). Figure 9.2-1 shows a typical result, which was specifically obtained from instance 1 of signal class B1P. The lost fraction for this example was 9.29%. The original signal is shown in Figure 9.2-1(a), while Figure 9.2-1(b) displays the approximation computed from the best set of curve fit parameters. The third graph, shown in Figure 9.2-1(c), is a plot of the residual signal. All three graphs are drawn to the same scale. It is clear that the approximation was very good, and that the largest discrepancies occurred at the beginning of the signal. This was to be expected because a sharp impact contained energy distributed over a wide range of high frequencies, and hence was not as well approximated by 5 terms as the later portion of the signal in which the high frequency transients had mostly decayed away.

A few comments are in order regarding the interpretation of the curve fit parameters. The curve fit function was a sum of 5 terms which were identical in form, each being determined by 4 independent parameters. Because of this, there was no obvious means of directly comparing two terms from two different signals. For example, suppose (as was the case) that B1P signals were observed to have slowly decaying components at 5106 and 3100 Hz. These two frequencies may have corresponded to the first and third damped sinusoids fitted to instance 9, and the second and fourth fitted to instance 14. In other words, the actual value of the function determined by the parameters (A_j, b_j, v_j, and ϕ_j, $1 \leq j \leq 5$) was not changed by exchanging two different values of the index j. The question then was in what order should the fit parameters be placed to permit comparisons between them.

Several different orderings of the terms were tried, in particular arranging them in order of descending amplitude, ascending frequency and ascending decay coefficient. The latter proved to be the most useful. It turned out that commonalities among different instances of the same signal class were readily apparent when the terms were arranged in this way. A plausible explanation of this fact can be made by considering the physics of the signal production. The largest cause of variability in the production of Air signals of the same class was unavoidable variation in the impact of the striker with the target. This had the largest effect on the initial shape of the signal, and hence on the transient (quickly decaying) components. After the initial impact, signals from a





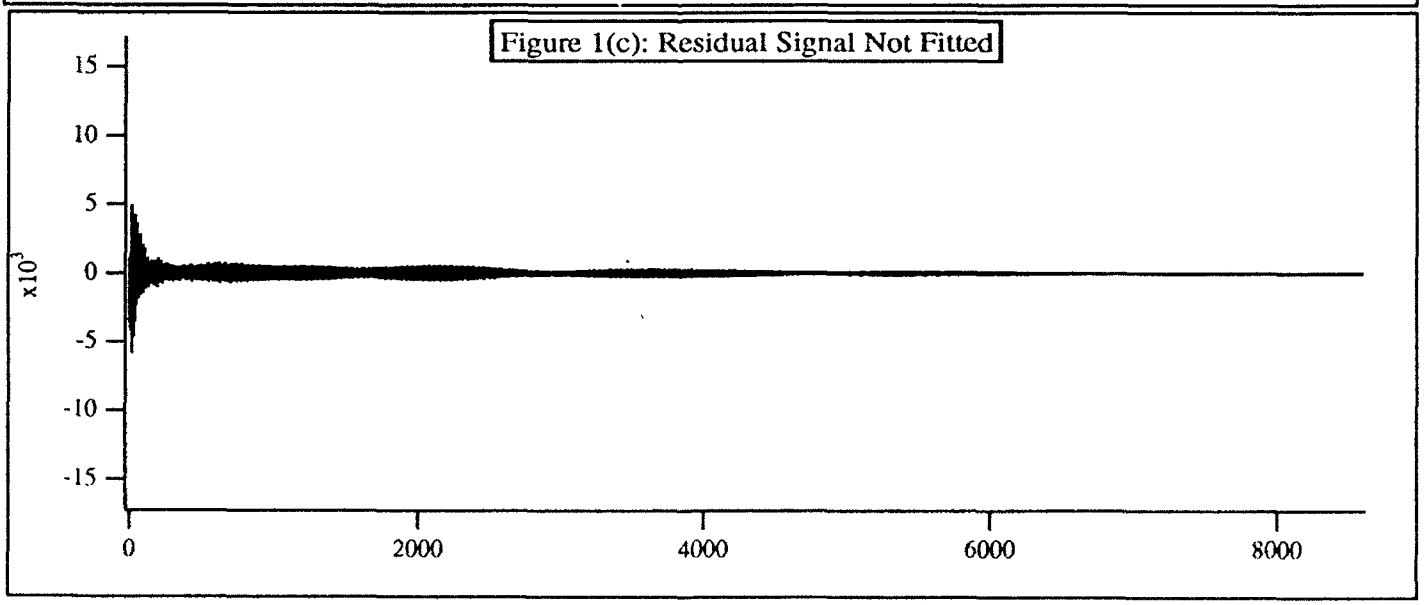


Figure 9.2-1 Curve Fit Approximation and Residual for Instance One of B1P

particular class were likely to have similar ring characteristics. It was therefore understandable that the terms with the smallest decay coefficient (longest ring) were similar, while terms with larger decay coefficients were more prone to variation.

In summary, time domain signals from the Air signal set were well approximated by a sum of five damped sinusoids whose parameters were obtained from standard chi-square minimization techniques. A method for determining the starting point for fits, and criteria for judging the fits were developed. The end result for each signal was a set of parameters which approximated the shape of the signal very well, even on a point by point basis. The strongest commonalities between the parameters for different signals of the same class were found in the most persistent (slowest decaying) modes present. When the terms were arranged from slowest to fastest decaying, meaningful correlations to hidden nodes, human subject behavior and signal statistics were observed, and will be described in more detail in Section 10.

9.3 CORRELATIONS

For all of the relevant measures for a particular signal class, the correlations between those measures and the values of the signals on the human scaling dimensions were computed. The signal parameters were computed on one instance of each class of each signal condition (free-field, bottom, and air). In the cases of Free-field and Bottom signals the values resulting from computing the parameters on different instances of the signals differed by vanishingly small amounts. The differences between the parameters computed on different instances of Air signals was somewhat higher, in keeping with the greater variability within a class of Air signals, but was still small relative to the variability across classes.

The correlations are used in the following section to identify strategies corresponding to the human dimensions. Parameters which are highly correlated with a human dimension may be related to the underlying signal feature or strategy of that dimension.

10.0 DIMENSION INTERPRETATIONS

The dimensions which resulted from the scaling algorithm run on the subject confusion data have been discussed above. Several methods of characterizing the original signals have also been introduced and applied to the signals. It remains to relate these methods and their results to the dimensions to create models of those dimensions. These models then suggest which signal features the subjects were using along each dimension.

10.1 ANALYSIS METHODS

Each analysis tool fit into the framework described below. The analysis of hidden nodes, which is less familiar to most readers, is described in greater detail.

10.1.1 Overview of Methodology

At this point we had developed several tools for the interpretation of the signal dimensions and the comparison to networks. We had the dimensions themselves and the associated subject weights, which were discussed previously. The subject weights provide information about the extent to which each subject used the various dimensions in the scaling solution. The signal statistics described in Section 9 were examined for correlations to the various dimensions. A high correlation was assumed to indicate that the subject was listening for a feature related to that statistic. For the Air signals only, the statistics included the curve-fit parameters. In addition to their use in correlations, the statistics were used to build regression models of the dimensions. This showed which statistical signal features were most useful in predicting the placement of signals on a dimension, another clue to the subjects' strategies. An additional important clue came from listening to the signals. While the features noticed during aural examination can only be described here, they were quite useful in guiding the investigation of the dimensions.

Finally the network nodes were examined. Many of the individual hidden nodes which make up the networks described earlier were highly correlated with signal dimensions. That is, the activation levels produced at the output of a node by signals of each class were highly correlated with the placement of those signals on a dimension. When a node was found to be highly correlated with a dimension, the node was examined in detail to determine its method of producing particular activation levels for the various signals. In some cases the node's strategy closely

matched the strategy derived via other analyses (such as the statistical models) of the dimension. In other cases the node's method suggested other means of reaching the same distribution of signals.

Certain hidden nodes, particularly those correlated with the first dimensions of the Air signals scaling solutions, are treated in greater detail than other nodes. The difference in depth illustrates the level of analysis possible without burdening the reader with the text associated with these analyses for all of the several hidden nodes.

The correlations between the scaling dimensions and these various tools and measures are shown in each case by a figure. The figures are an aid to understanding the relationships between the dimensions and the correlated signal statistics and hidden nodes. Using

H_0: $\rho = 0$

H_1: $\rho \neq 0$

$N = 12, \ \propto = 0.01, z_{.005} = 2.575$

$z = ((n-3)^{1/2} / 2) * \ln((1+r)(1-\rho) / (1-r)(1+\rho))$

or $r = 0.6954$

suggests that a 0.70 absolute correlation is significant at the 1% level. Therefore, the dimensions figures show absolute correlations of 0.70 or higher, except when a correlation close to 0.70 is included for parallelism to another dimension.

10.1.2 Analysis of Specific Hidden Nodes

The analysis of the functional roles of a given hidden node will be completed in three stages. The starting point will be an examination of the weights connecting the hidden layer to the output layer. By comparing the weight given the hidden node in question to the weights placed on other hidden nodes, it is possible to determine the purpose for which that hidden node is used. With this information in mind, the weights between the input layer and that hidden node will then be

explored to determine what information in the signal the hidden node uses to perform its function. The picture is completed by evaluating the response of the node to actual signal inputs. Once a hidden node is analyzed, it may be compared to others to gain insight into the behavior of the networks as a whole.

Before addressing the physiology of specific hidden nodes, a general discussion of the output layer will be helpful. To facilitate the discussion, terms appropriate to Air networks will be used as necessary (e.g. Plastic Striker). Unless otherwise specified, however, the comments are general and may be applied to Free-field and Bottom networks with suitable substitutions for terms specific to the Air signals (e.g. Angle for Striker).

The output layer divided naturally into three groups: the Material nodes (B and S), the Thickness nodes (Ten and Five) and the Striker nodes (M, P, and W). Within the Material and Thickness nodes, the binary nature of the classification performed resulted in some simplification. Because the target output of output node B was always 0.0 whenever S was 1.0 and vice versa, the output nodes B and S consistently developed (nearly perfectly) equal and opposite connections to the hidden layer. The same is true of the Thickness output nodes (for examples, see Figures 10.3.1-1 and 10.3.1-2).

Although the classification of Striker involves placing the signal in one of three categories, similar relationships sometimes evolved between two of the three Striker output nodes. When present, this "pseudo-binary" structure may imply that the network learned to recognize only two of the three Striker types, with the third being recognized by default. These relationships were never as perfectly equal and opposite as those which occurred in the inherently binary classifications of Material and Thickness. For example, in Figure 10.3.1-2, the weights found by this network's Metal output node, M, are of opposite sign, but much larger in magnitude than those of the Wood output node, W. A relationship nevertheless exists; for each of these output nodes, the relative importance of each hidden node is approximately the same. The same hidden node activations which activate one node will tend to suppress the other.

10.2 DIMENSIONS OF THE BOTTOM SIGNALS

The relationships among the first two scaling dimensions of each Bottom scaling solution and the related signal statistics and hidden node activations are shown in Figure 10.2-1.

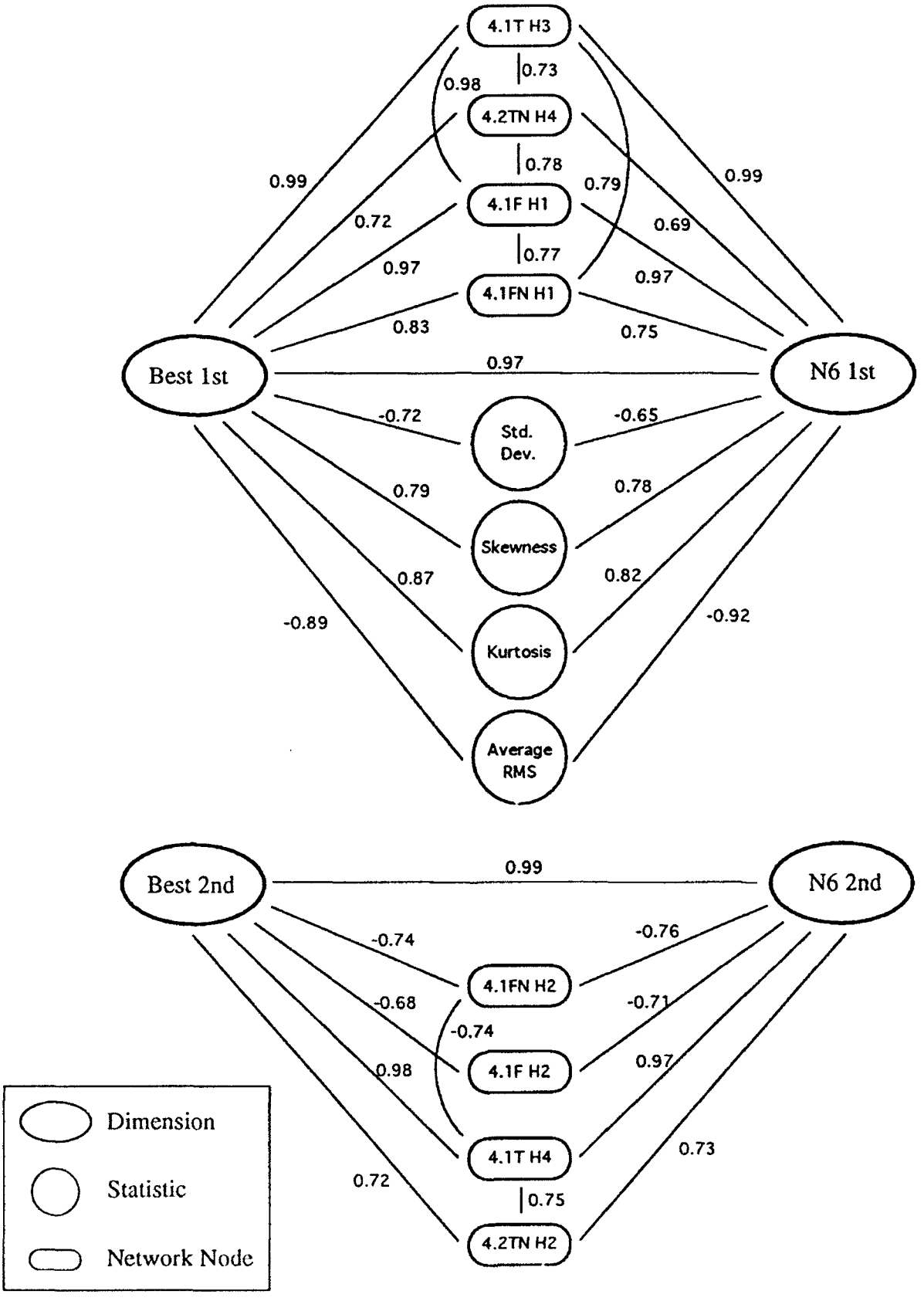


Figure 10.2-1 Correlations Among Bottom Scaling Dimensions, Statistics, and Network Hidden Nodes.

The scaling solutions for the two Bottom cases ("Best" and "N6") are extremely similar in the first two dimensions. The first two pairs of dimensions are correlated at 0.97 and 0.99, respectively. It appears that subject N6 applied the same strategy as did the three subjects as a group. This makes sense in light of the difficulty the subjects had with the Bottom signals, and the apparent high importance of the large reflection from the 90° objects in comparison to any other feature in that or other orientations. The subjects had relatively little information to work with, and the information present was almost completely defined by the 90° reflection. However, subject N6, as well as one other subject who is included in the Best solution, could make discriminations among the three orientations beyond just identifying the 90° signals. That is, they could also classify 0° and 45° signals, as shown by Angle test scores of 94% and 96%. This capability is reflected in the second dimension. Since this capability is rare among the subjects, defining it was of increased importance.

Weirdness values for both Bottom scaling solutions indicate that dimension one was much more important to the subjects than any other. This data also fits the theory that the 90° reflection dominated any other features. These subjects were selected for their high scores, which are due primarily to high performance on the angle parameter. The selection of these subjects probably led to the importance of the second Bottom dimension in each solution. In each case the second dimension has a weirdness score of approximately one-half the first dimension. This indicates that the second dimension is of significant importance; when the weirdness information is combined with the breakdown of signals by Angle, the second dimension attracts particular interest. The third dimension, however, is of such little importance in the subject's classification that it is not modeled here.

10.2.1 First Dimensions for Best and N6 Scaling Solutions

As seen in Figures 7.4.2-1 and 7.4.2-2, the first dimensions of each scaling solution are very similar and serve to discriminate the 90° signals from the other two orientations. The 45 and 0° signals are placed very close to one another, while the group of 90° signals is some distance away. Only in the Best first dimension do we see a slight variation, in which S59 is slightly lower than the cluster of other 90° signals.

10.2.1.1 Dimensions Analysis

Listening to the signals in the order found on this dimension strongly suggests that the subjects are making the postulated distinction between 90° signals and the other two angles. Due to their orientation broadside to the insonifying wave, the 90° signals contain a reflection from the target which is relatively large compared to the bottom reflection. The reflection is clearly audible in the 90° signals, and absent in the others. As a signal feature this reflection dominates any others that the casual listener is likely to find, leading to the heavy reliance on the first dimension shown in the scaling results.

Although the casual listener is impressed with the 90° reflection in the time domain, both first dimensions are correlated with three statistics in the frequency domain: standard deviation, skewness, and kurtosis. These are all descriptions of the shape of the distribution of frequencies in the signals. For instance, 90° signals have a smaller standard deviation according to that correlation, indicating a narrower band of frequencies, than 45° and 0° signals. They also seem to be more skewed than 45° or 0° signals. The important point is that the easily recognized time domain feature is reflected in the frequency domain as well. The regressions described below use these frequency domain statistics as well. The preservation of this feature in some form across the transform from time to frequency domains also helps explain how the neural networks can find information from the frequency domain input to classify the Bottom signals. Such information is actually present to be used in classification, in addition to artifactual information which networks may learn to employ.

As a time domain measure the root mean squared (RMS) level of the first and ninth instances of each class was computed, and the two were averaged for a representative measure of the class. The average RMS level is highly negatively correlated with the Best first dimension and with the first dimension of N6. This is likely to be due to the preprocessing of the signals. The maximum level of all signals was equalized. This makes the bulk of the 90° signals lower in amplitude than equivalent portions of the 45° and 0° signals. This difference is reflected in lower RMS values of the 90° signals. Note that, because the 90° reflection is so large, it would stand out in any RMS measurement. Had the signals been equalized to the bottom reflection, the 90° signals would have had higher RMS values than 45° and 0° signals.

One may adequately predict the values of the signals on the first of the Best solution by a regression equation using only the average RMS:

$R^2(adj) = 77.9\%$
$p \leq 0.0000$

When frequency domain measures are used in the regression, a slightly better set of predictors is found:

$R^2(adj) = 83.0\%$
Kurtosis $p \leq 0.0000$
Low Frequency Slope $p = 0.0263$

Regression models for the N6 solution are very similar. Average RMS by itself produces:

$R^2(adj) = 83.9\%$
$p \leq 0.0000$

While the same set of frequency domain predictors give:

$R^2(adj) = 80.1\%$
Kurtosis $p = 0.0001$
Low Frequency Slope $p = 0.0143$

While both time and frequency domain parameters make good regression predictors for both first dimensions, they do not combine to make a better predictor. This indicates that the information in them is redundant as regards the first dimension. This makes sense if the time domain event of interest, the 90° reflection, produced the frequency domain differences demonstrated by the regressions and correlations.

10.2.1.2 Analysis of Bot4H(1)T-H3 and Bot4H(1)TN-H4

The hidden nodes from Bot4H(1)T and Bot4H(1)TN will be referred to here as T-H3 and TN-H4. T-H3 was almost perfectly correlated with both dimensions, while TN-H4 was correlated at 0.72

and 0.69 with the Best first and N6 first dimensions respectively. The output layer of Bo:4H(1)T, shown in Figure 10.2.1.2-1, indicates that the only role of T-H3 was to detect 90° signals (and it is the only means of doing so). This was well in keeping with the division of the signals on both dimensions. High activation from the node is used to activate the 90° output node as well as to suppress the other angle output nodes. TN-H4, in contrast, has roles in the Material and Thickness outputs as well as Angle (see Figure 10.2.1.2-2). Within Angle TN-H4 serves to detect 90° signals and is the only means of doing so. It suppressed 45° output but not the 0° output, a significant difference from T-H3.

The input weights of T-H3 are shown in Figure 10.2.1.2-3. There are two groups of weights: I1 to I15, generally positive and including the large weights on I13, I14, and I15; and I16 to I43, almost all negative and significant. 90° signals all have their dominant energy in I13 - I15. An example is shown in Figure 10.2.1.2-4(a). 90° signals are detected by the large weights on these bins. Energy drops off rapidly in all 90° signals after these bins, so the large negative weights at higher frequency bins have little effect on 90° signals. 45° and 0° signals have most of their energy after I15, and are rejected by the large negative weights in the range I16 - I43, as shown in Figure 10.2.1.2-4(b). Figure 10.2.1.2-5 shows the final activations of all classes, with only the 90° signals activating the node.

The input weights of TN-H4, shown in Figure 10.2.1.2-6, are more complex than those of T-H3. This is unusual in that weights from networks trained with noisy inputs are generally simpler than weights from networks not trained with noisy inputs. The output layer of the parent network on TN-H4, described above, indicates that this node is being used for more functions than simply telling 90° signals from other angles, which accounts for a more complex weight structure. The output activations of TN-H4, seen in Figure 10.2.1.2-7, show that both 5% 0° signals receive high activation along with the 90° signals. these two signals are identified by the node by their high energy in bin 29. This is in keeping with the role of TN-H4 with respect to the Thickness output, where contributes to activating 5% and suppressing 10%. TN-H4 does not suppress the 0° output node, in keeping with the high activations for the 0° 5% signals. In order to implement this more complex strategy, TN-H4 needed a more complex weight structure than T-H3.

Figure 1(a)

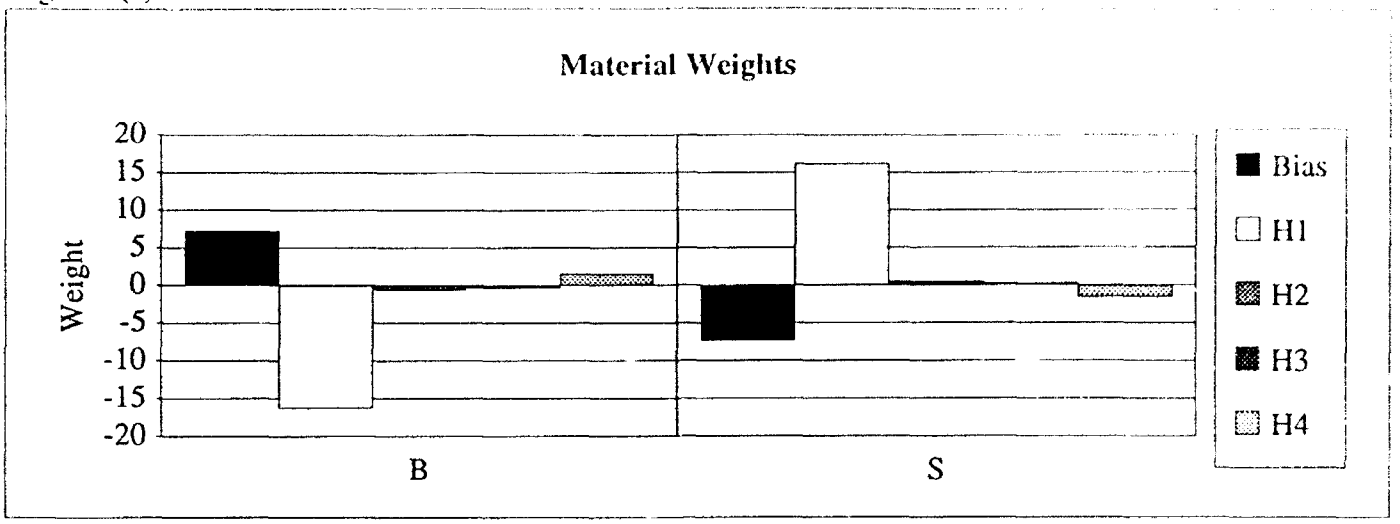


Figure 1(b)

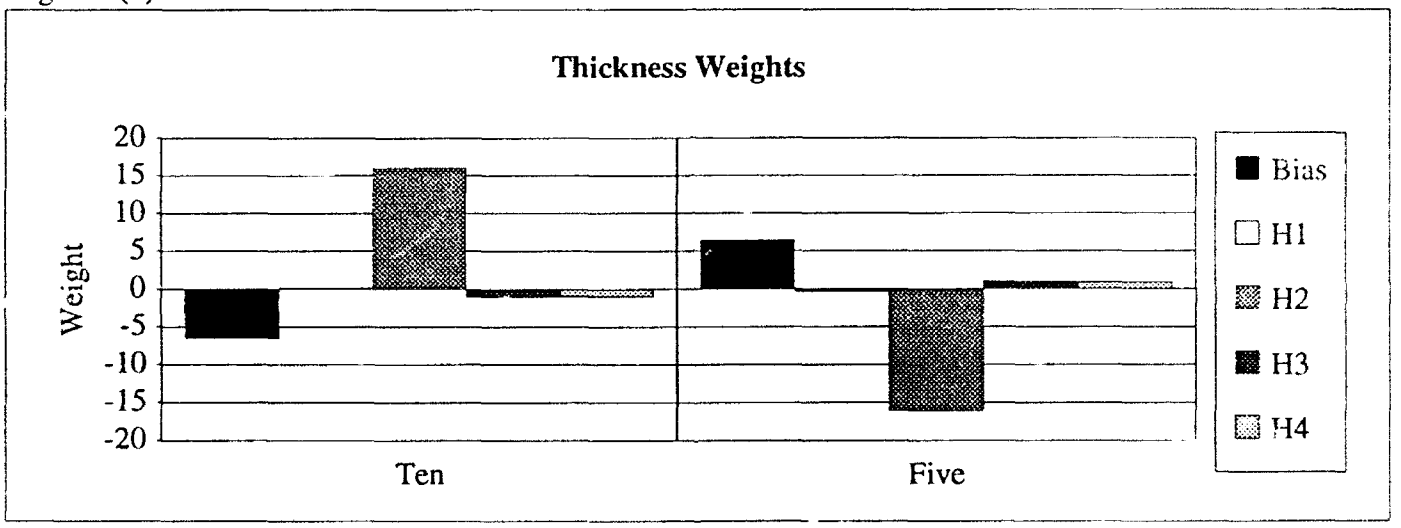


Figure 1(c)

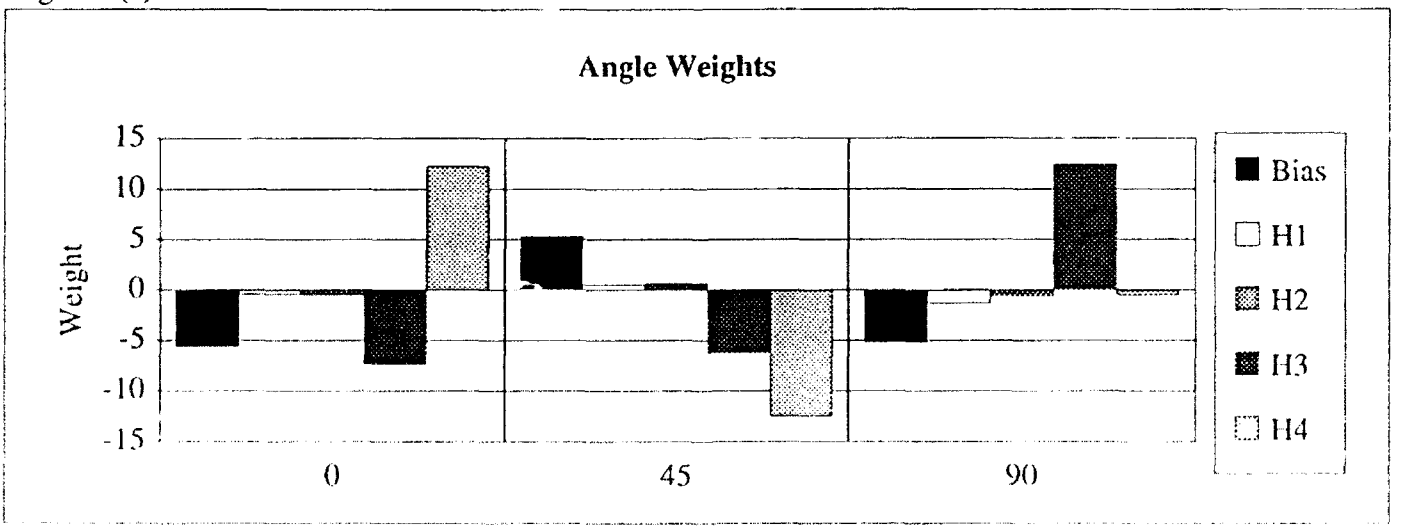


Figure 10.2.1.2-1 Weights on Hidden to Output Layer Connections in Bot4H(1)T

Figure 2(a)

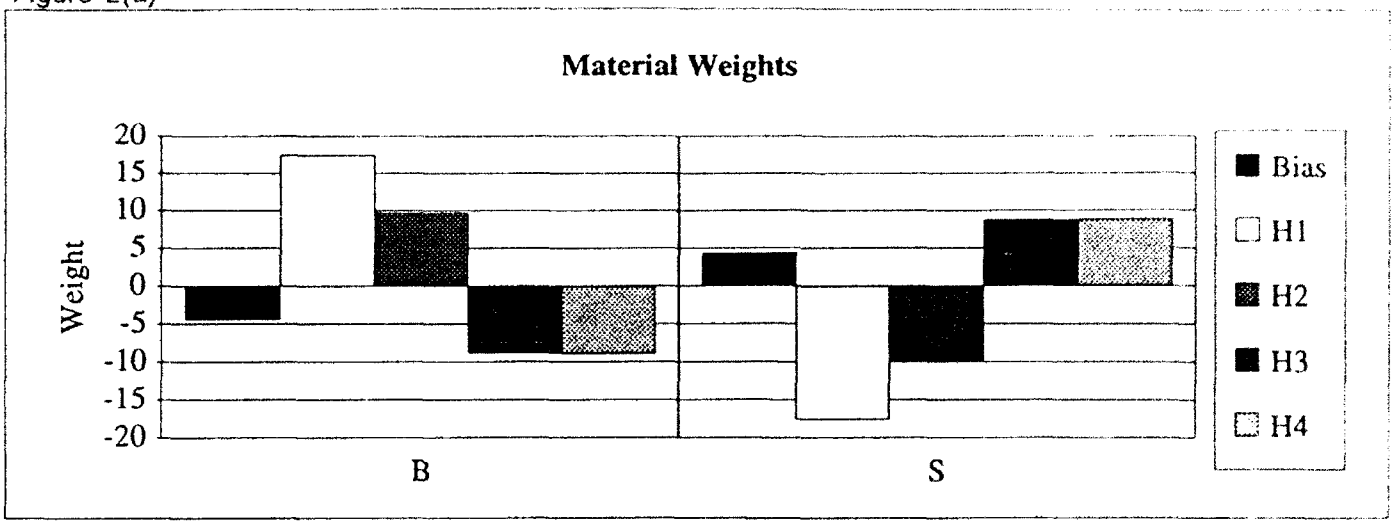


Figure 2(b)

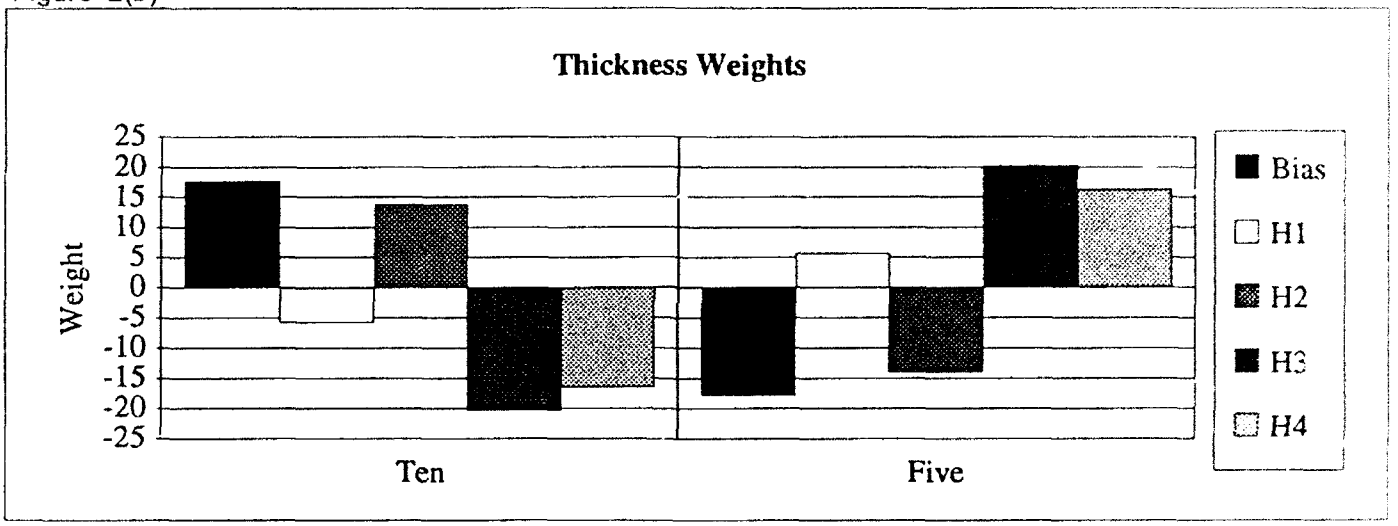


Figure 2(c)

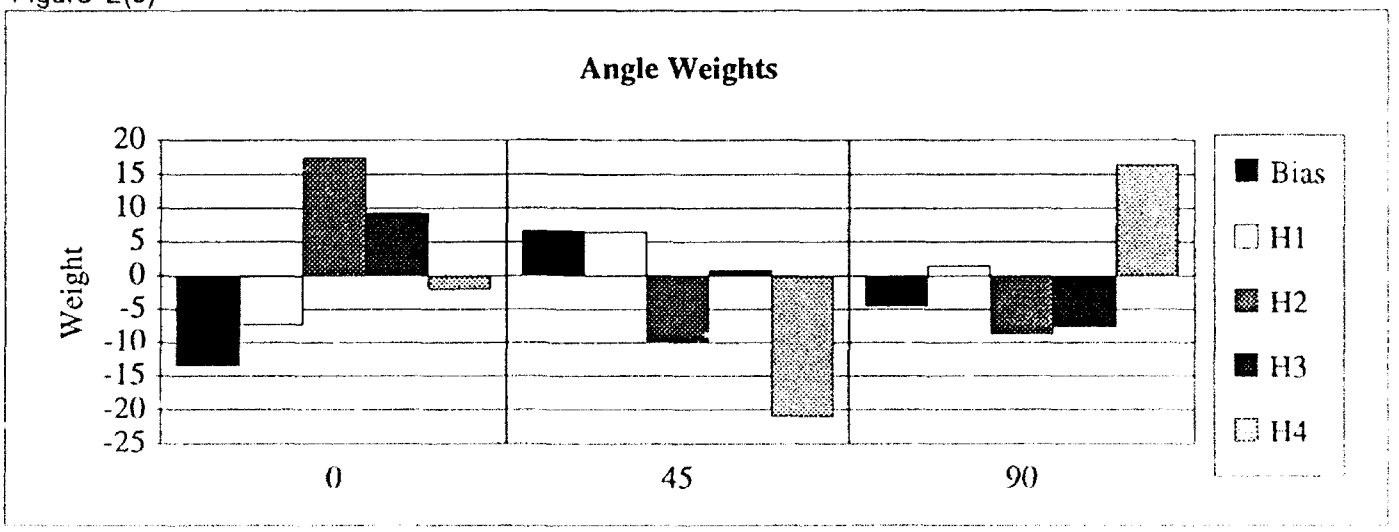


Figure 10.2.1.2-2 Weights on Hidden to Output Layer Connections in Bot4H(1)TN

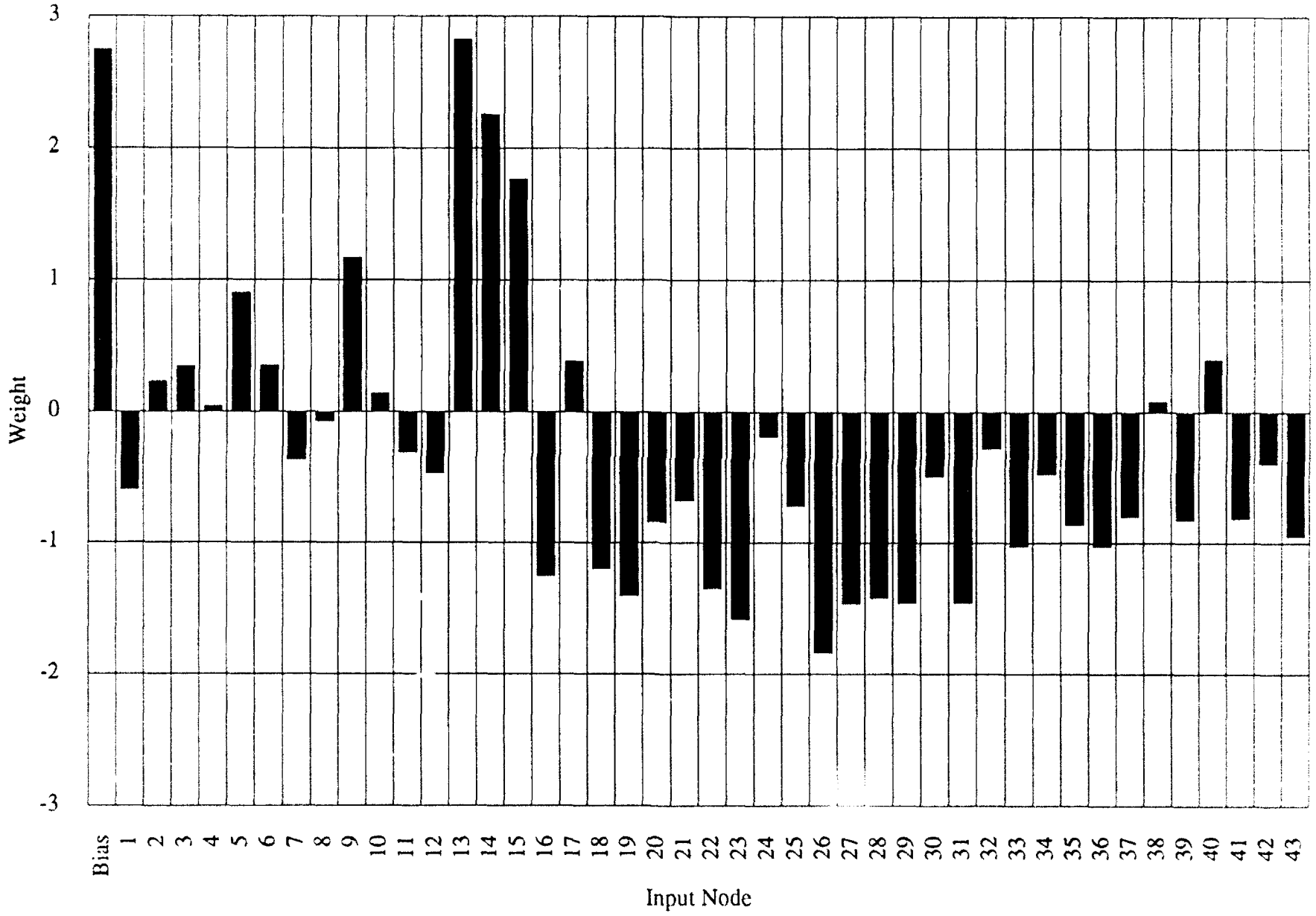


Figure 10.2.1.2-3 Weights on Input Layer to Hidden Node 3 Connections in Bot4H(1)T

Figure 4(a): S19

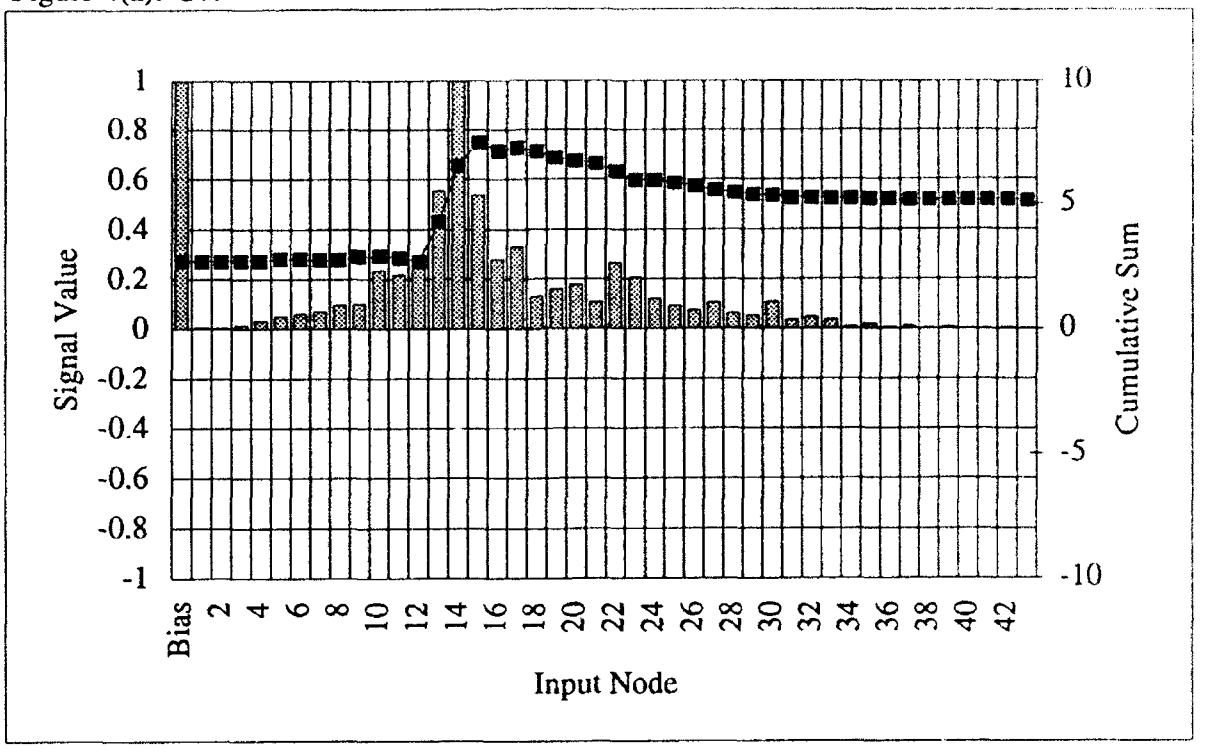


Figure 4(b): S10

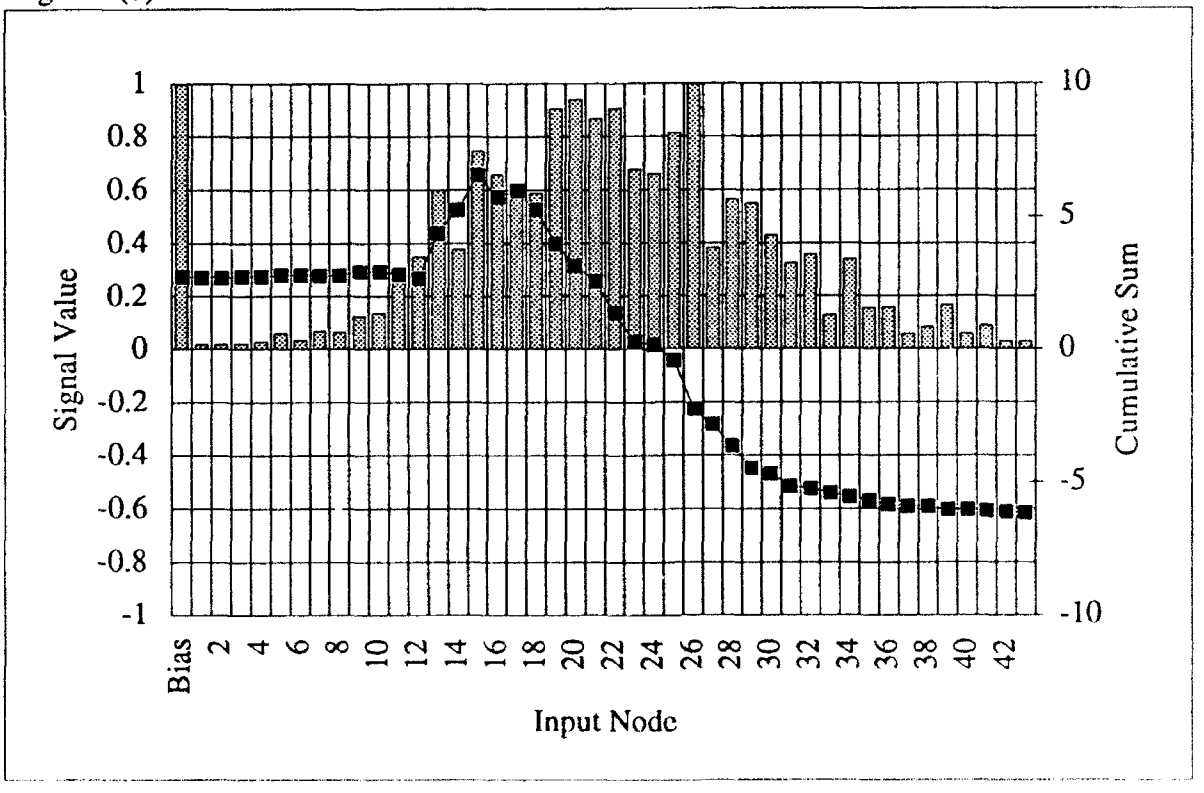


Figure 10.2.1.2-4 Cumulative Sum of Hidden Node Bot4H(1)T-H3 for Instance Nine of Classes S19 and S10

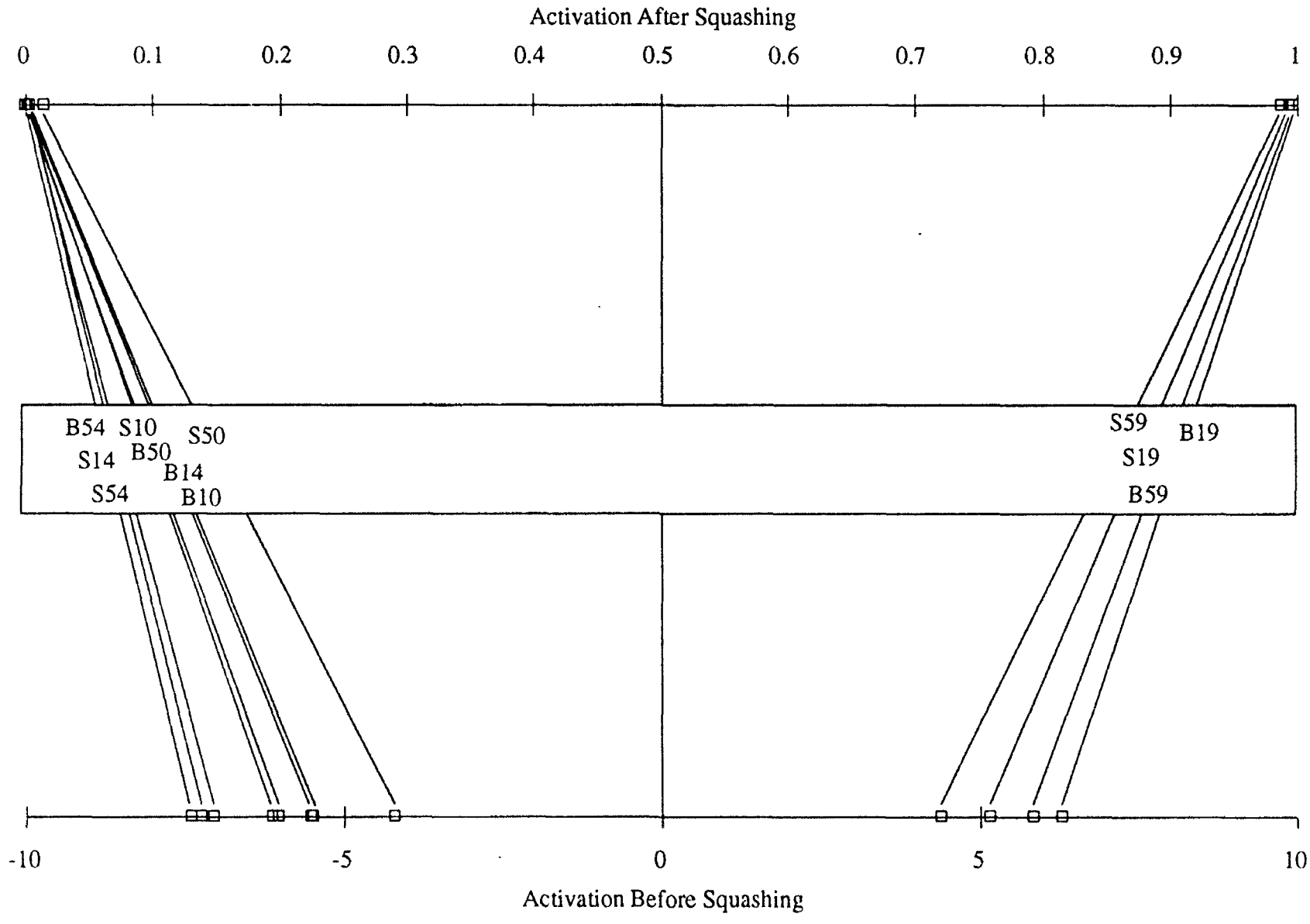


Figure 10.2.1.2-5 Hidden Node Bot4H(1)T-H3 Activation for Instance Nine of Each Signal Class

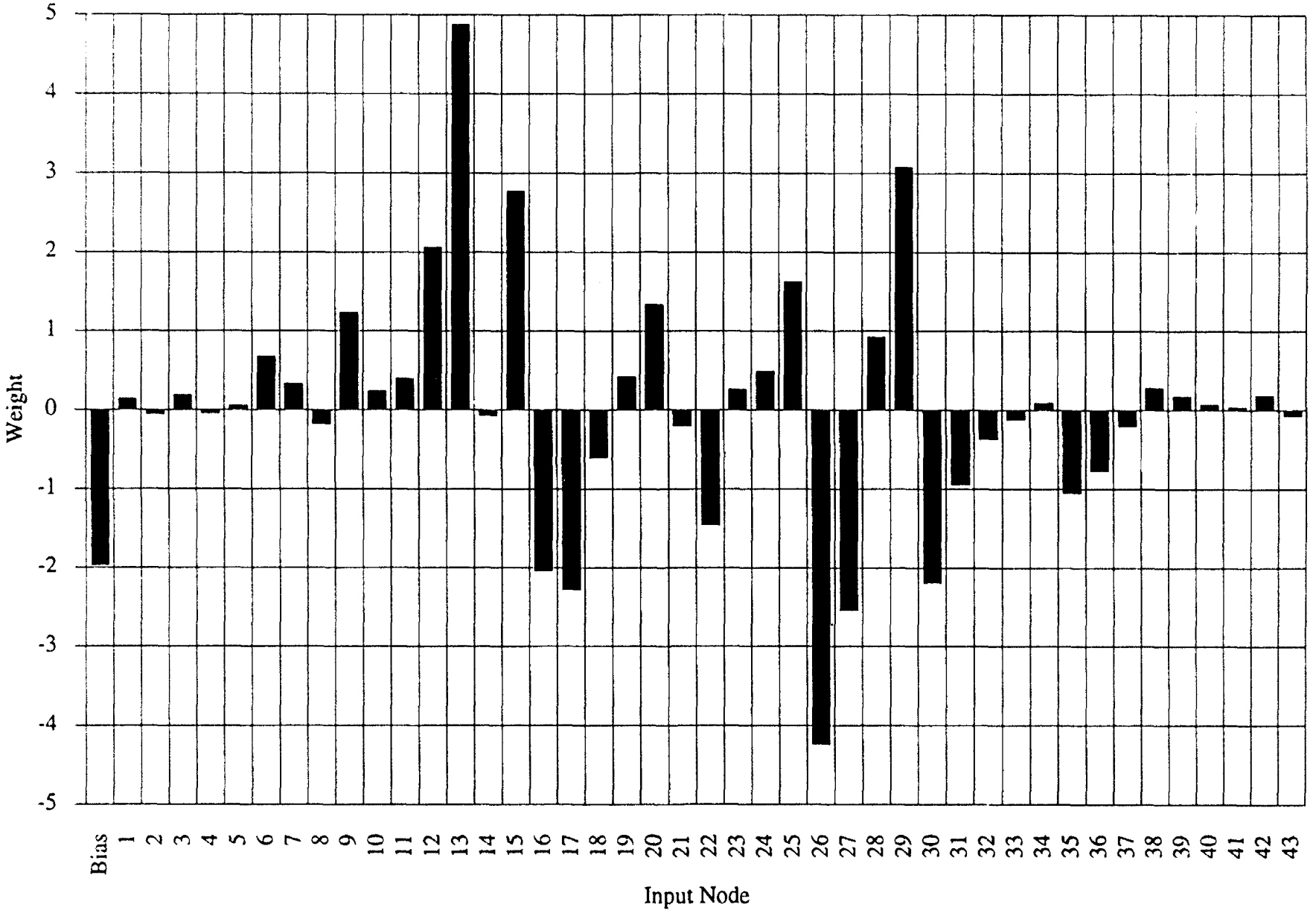


Figure 10.2.1.2-6 Weights on Input Layer to Hidden Node 4 Connections in Bot4H(1)TN

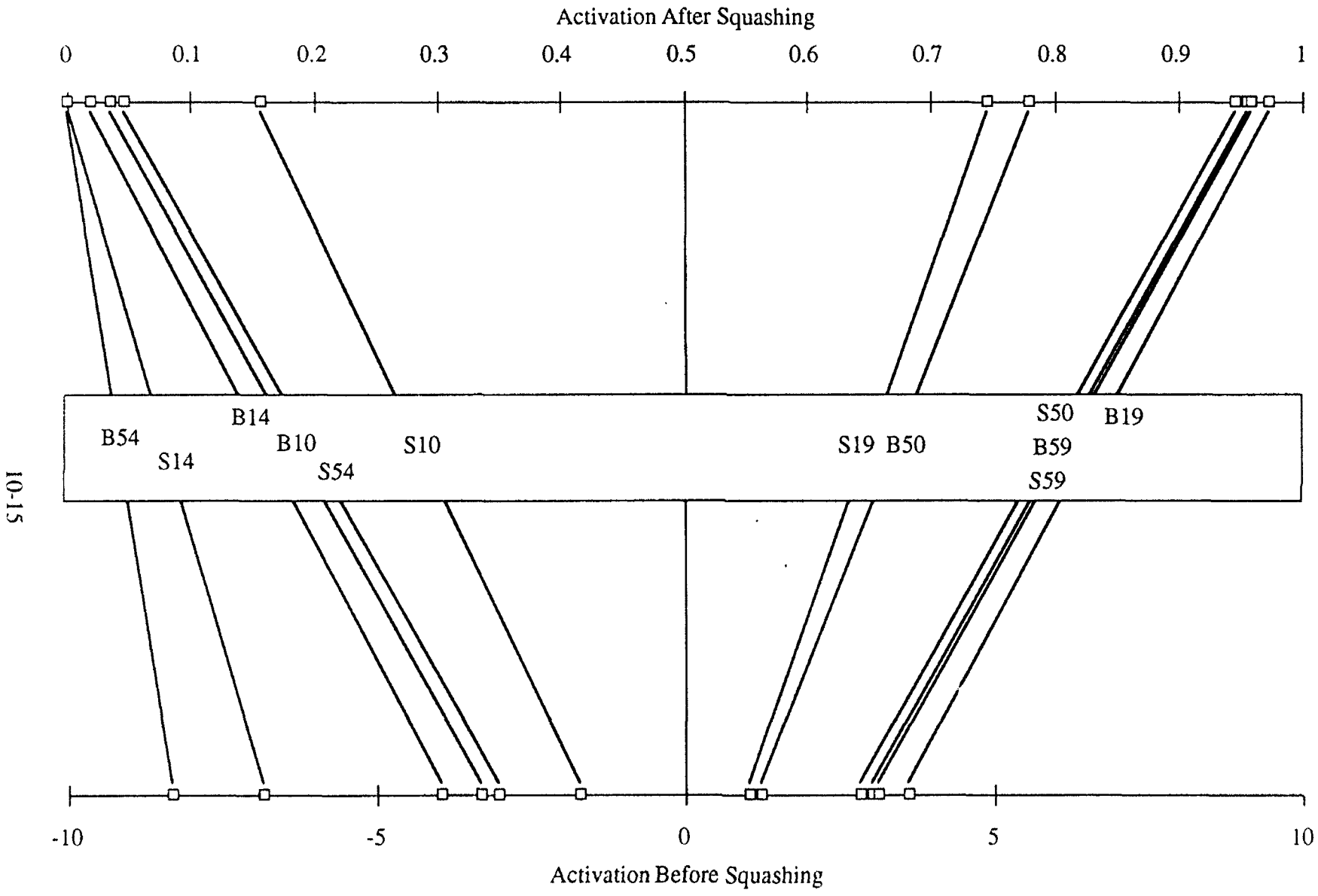


Figure 10.2.1.2-7 Hidden Node Bot4H(1)TN-H4 Activation for Instance Nine of Each Signal Class

10.2.1.3 Analysis of Bot4H(1)F-H1 and Bot4H(1)FN-H1

These nodes are referred to as F-H1 and FN-H1. Just as several frequency domain signal measures were correlated with the dimension, neural network nodes are able to extract information in the frequency domain to produce activations correlated with the dimensions. F-H1 is very highly correlated with both dimensions at 0.97, while FN-H1 is correlated with the Best first dimension at 0.83 and the N6 first dimension at 0.75.

Weights on the output layer of Bot4H(1)F, seen in Figure 10.2.1.3-1, indicate that the sole purpose of F-H1 is to detect 90° signals. It is used to activate the 90° output node, suppress both 45° and 0° output nodes, and is not used by Material or Thickness nodes. Bot4H(1)FN shows a more complicated role for FN-H1 in Figure 10.2.1.3-2. It is used to detect 90° signals, and to reject 0° signals, but contributes to the activation of 45° signals as well. It is also used to detect Steel and 10% signals.

The activations of F-H1 are shown in Figure 10.2.1.3-3 and confirm the node's role as detector of 90° signals. The input weights of F-H1, seen in Figure 10.2.1.3-4, are not particularly informative in isolation. Clearly bin 11 may play a strong role in detecting 90° signals, and this bin corresponds to the 400 kHz insonifying frequency. Bin 16 is likely to play a role in rejecting 0° and 45° signals.

When the 90° signals are applied to the node the cumulative activations, an example of which is shown in Figure 10.2.1.3-5, demonstrate the importance of the large weight on bin 11. Although one of the 90° signals peaks in bin 10 and one in bin 12, the product at bin 11 is always the largest contributor to activation. 45° and 0° signals are rejected by bins 8-10, 12, and 16, as seen in Figure 10.2.1.3-6. The lack of energy at bin 11 was important to rejecting 45° and 0° signals, and illustrated the relationship between this node's processing and the high correlation between the dimensions and the skewness measure. 45° and 0° signals tend to have relatively little energy at bin 11, instead spreading their energy to adjacent frequencies, resulting in higher standard deviations of the frequencies in the signal.

FN-H1 applied a different strategy towards a similar end, as seen in the activations shown in Figure 10.2.1.3-7. The 10% 45° signals received high activation along with the 90° signals. As we learned above the 45° output weight is moderately activated by FN-H1. The strategy of the

Figure 1(a)

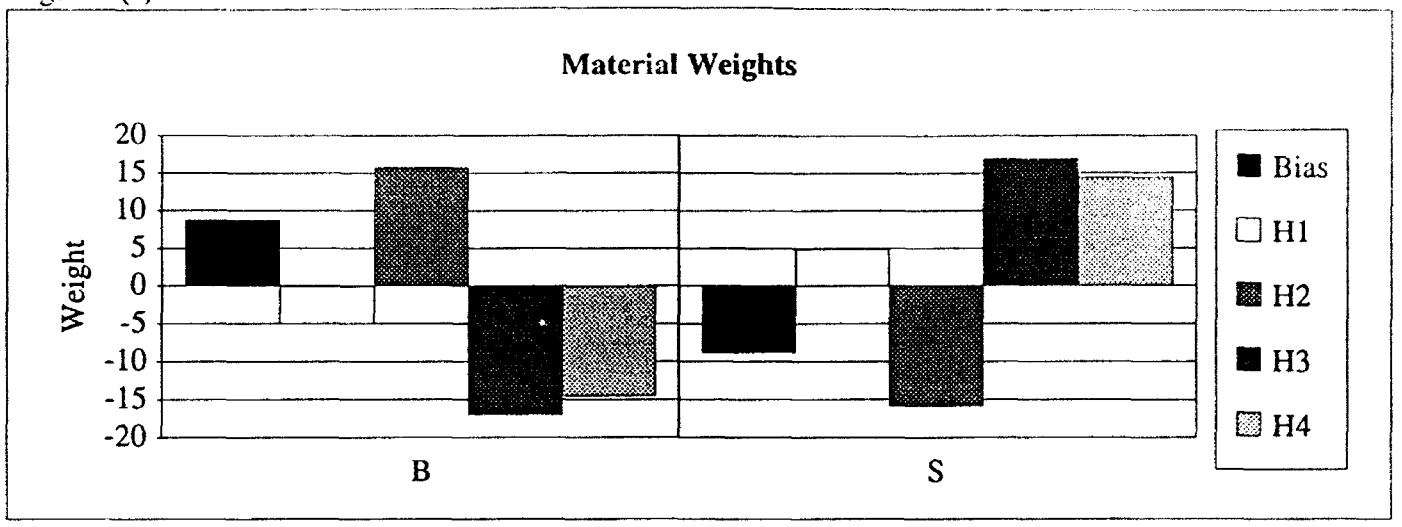


Figure 1(b)

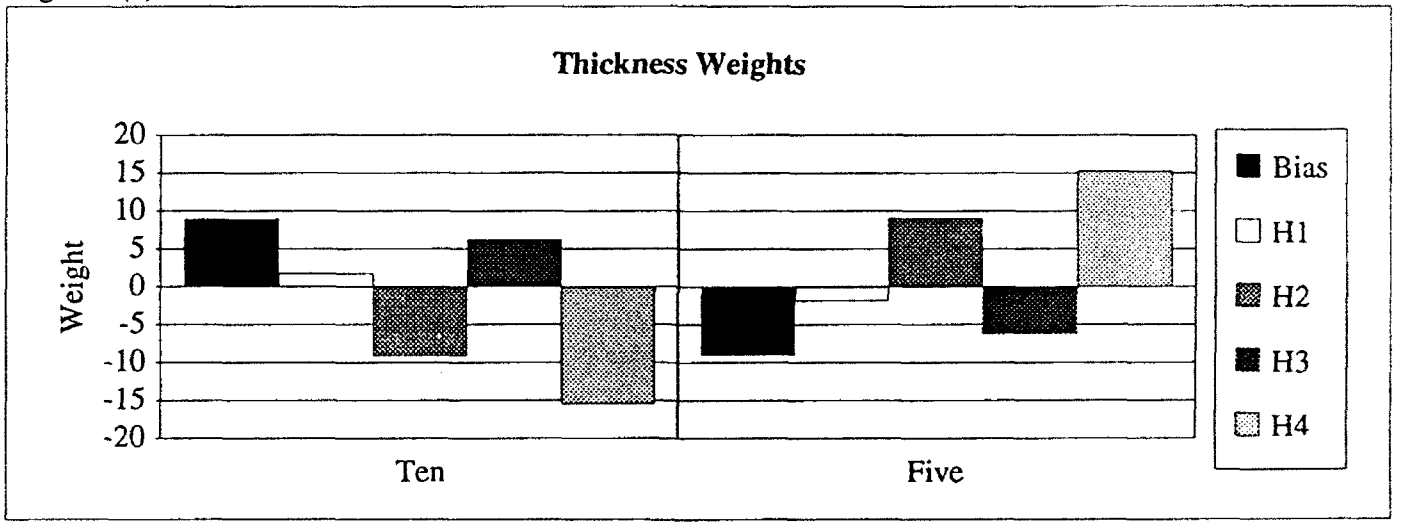


Figure 1(c)

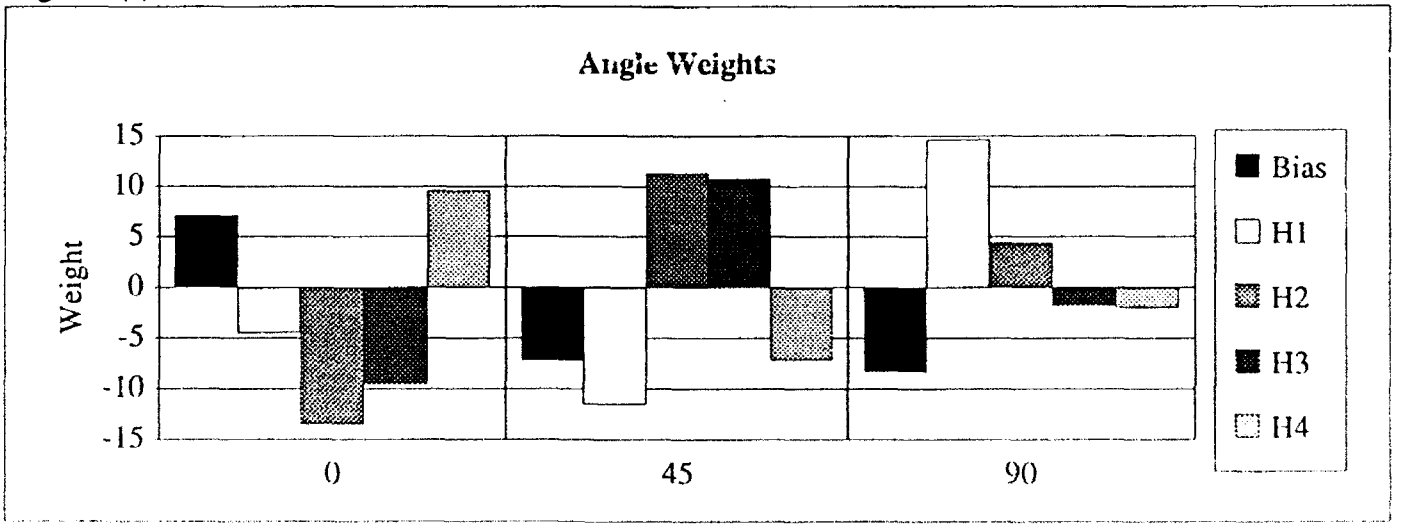


Figure 10.2.1.3-1 Weights on Hidden to Output Layer Connections in Bot4H(1)F

Figure 2(a)

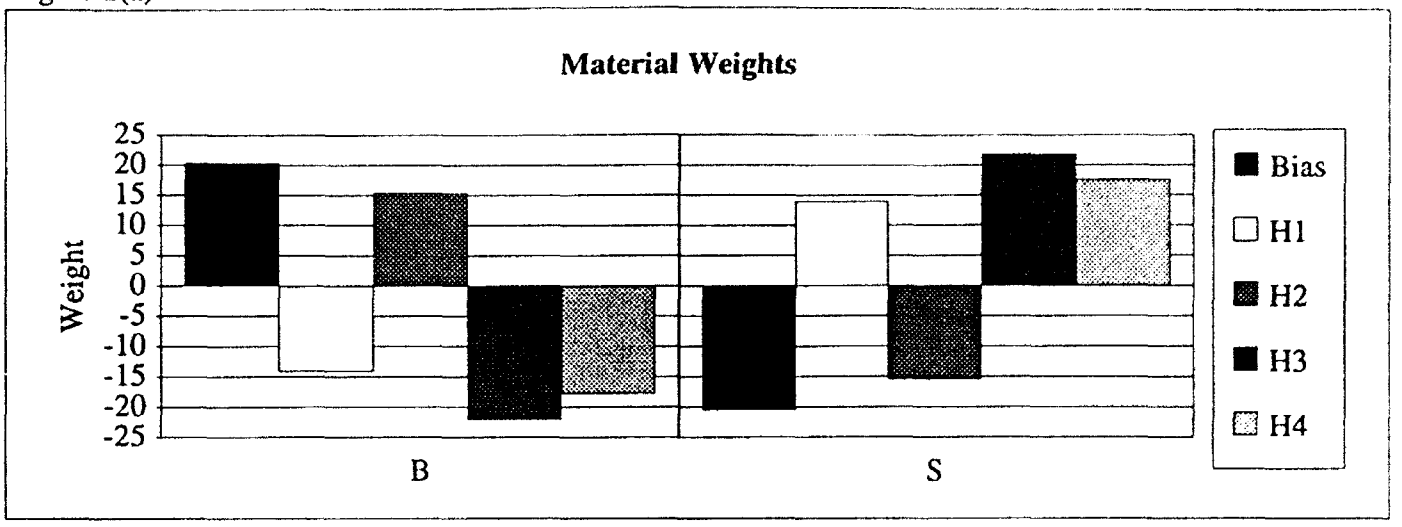


Figure 2(b)

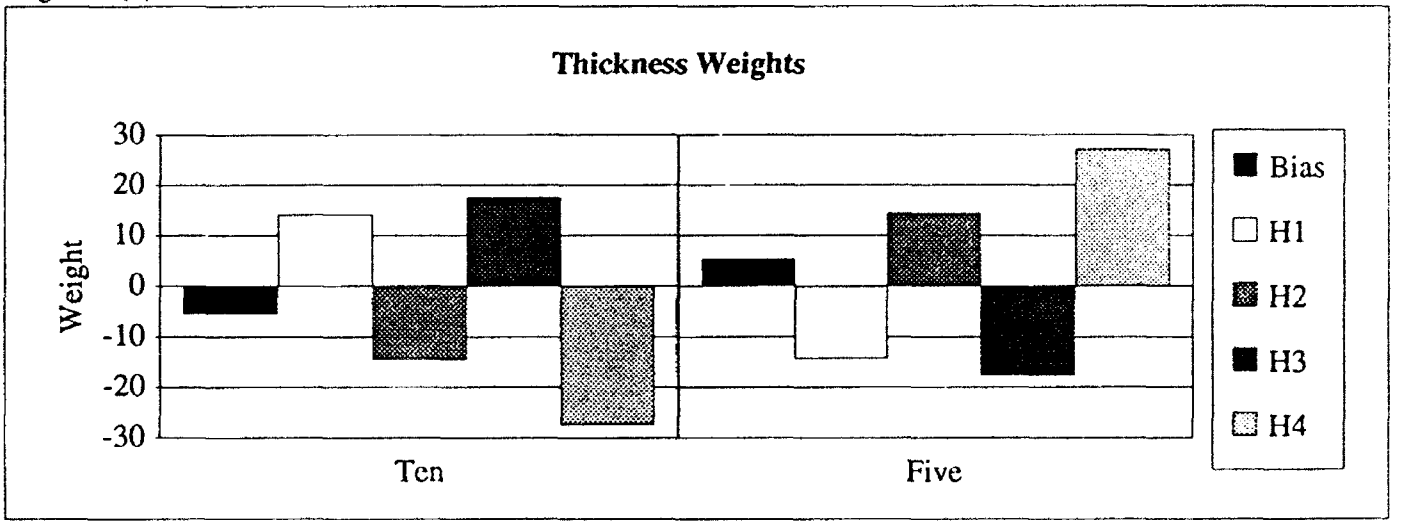


Figure 2(c)

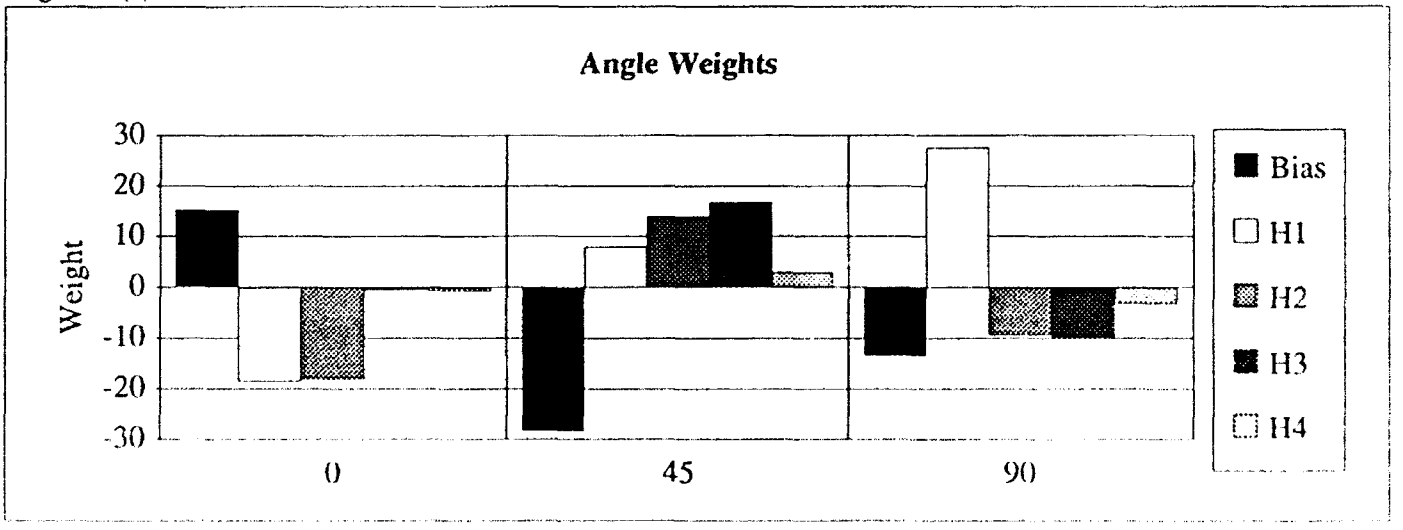


Figure 10.2.1.3-2 Weights on Hidden to Output Layer Connections in Bot4H(1)FN

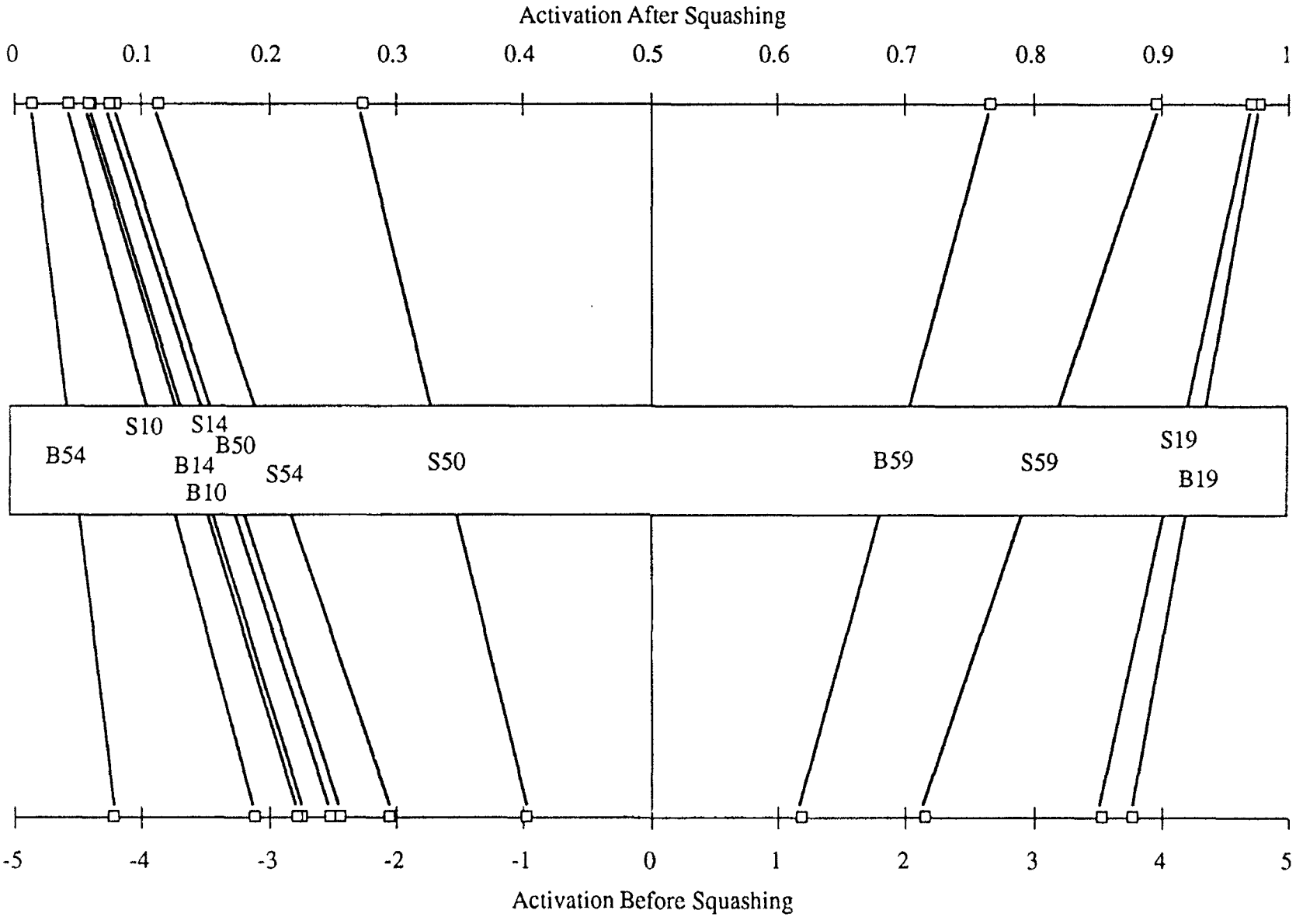


Figure 10.2.1.3-3 Hidden Node Bot4H(1)F-H1 Activation for Instance Nine of Each Signal Class

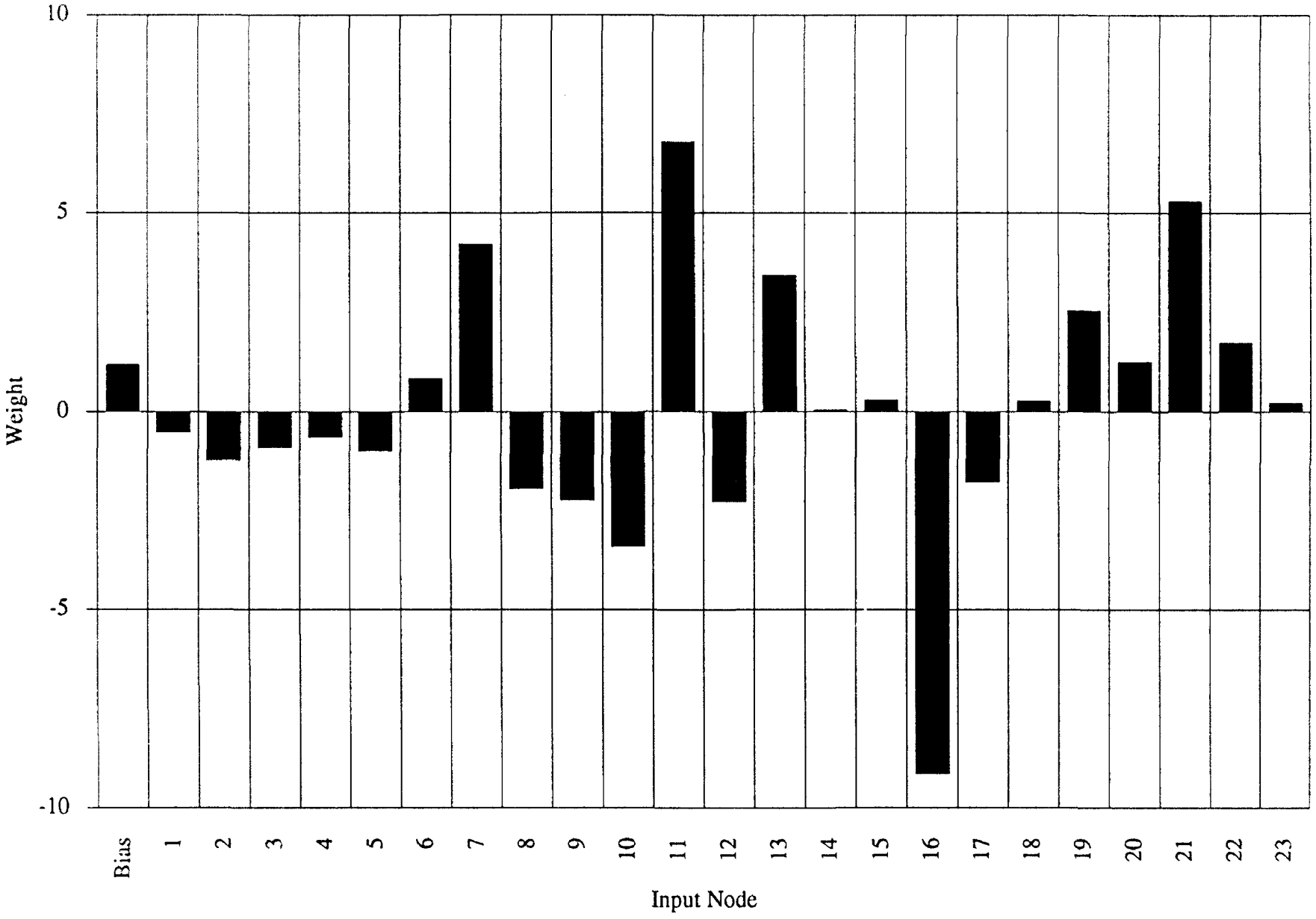


Figure 10.2.1.3-4 Weights on Input Layer to Hidden Node 1 Connections in Bot4H(1)F

Figure 5(a): B19

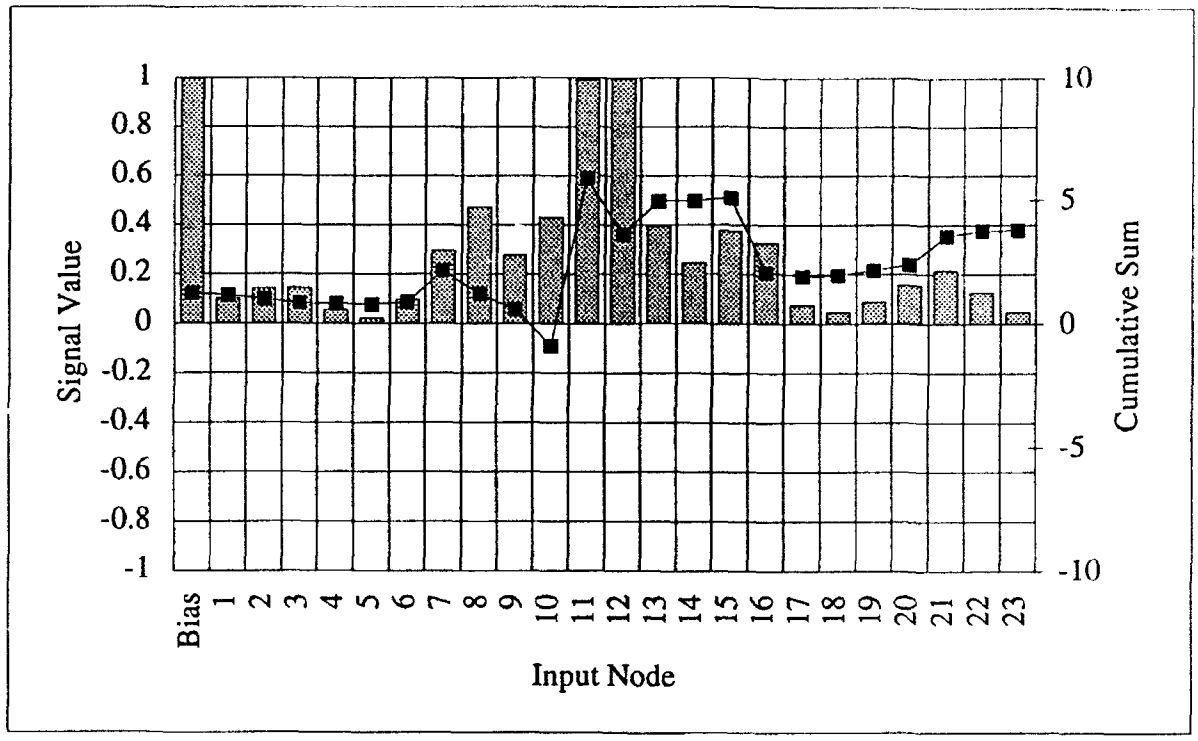


Figure 5(b): S19

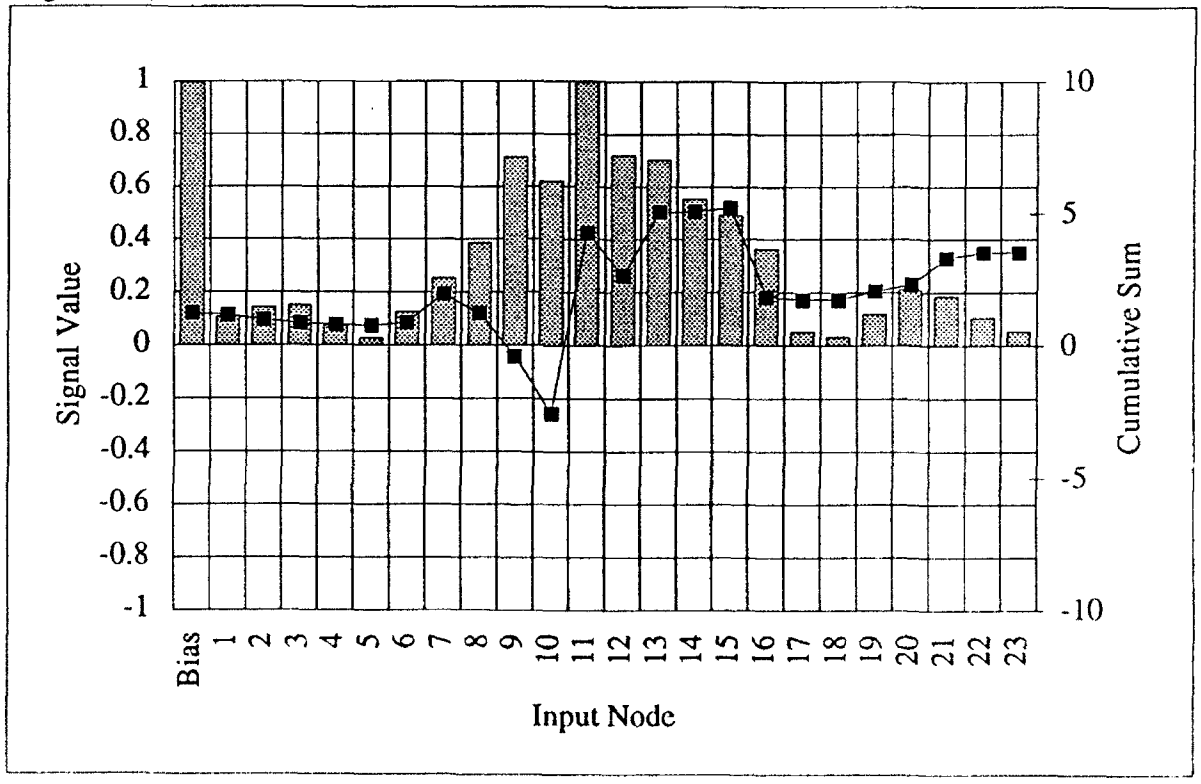
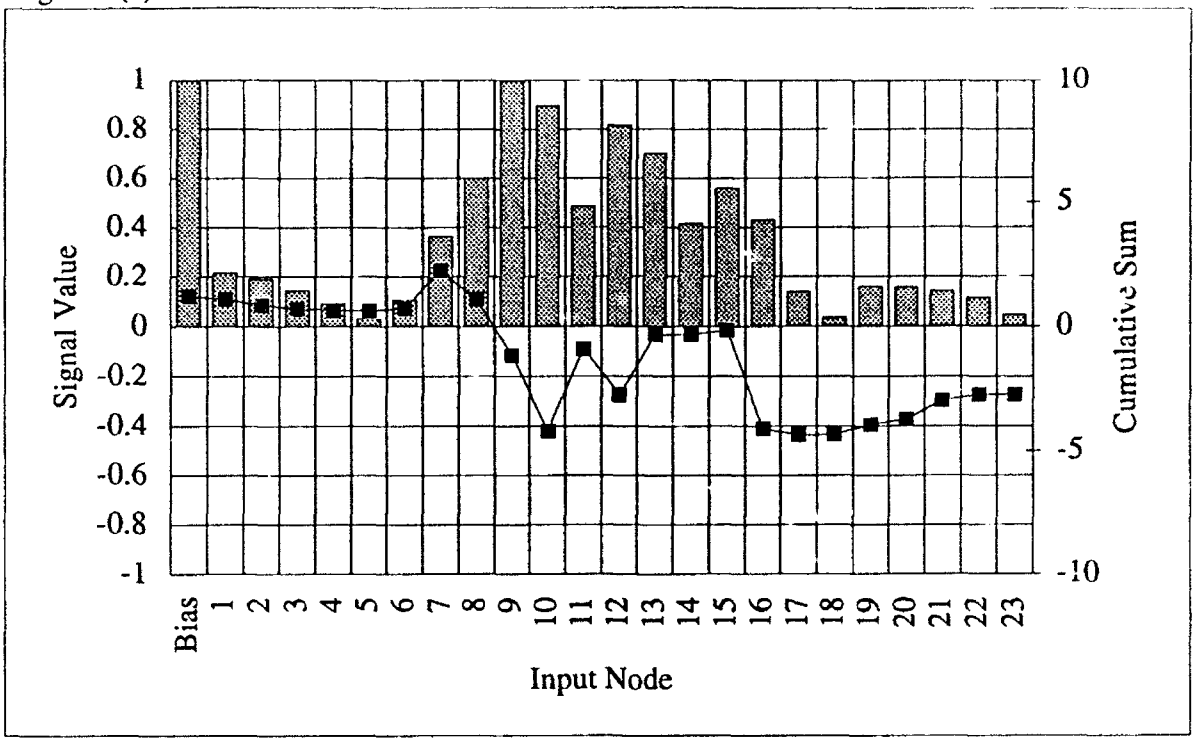


Figure 10.2.1.3-5 Cumulative Sum of Hidden Node Bot4H(1)F-H1
for Instance Nine of Classes B19 and S19

Figure 6(a): B10



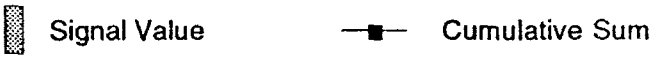


Figure 6(b): S10

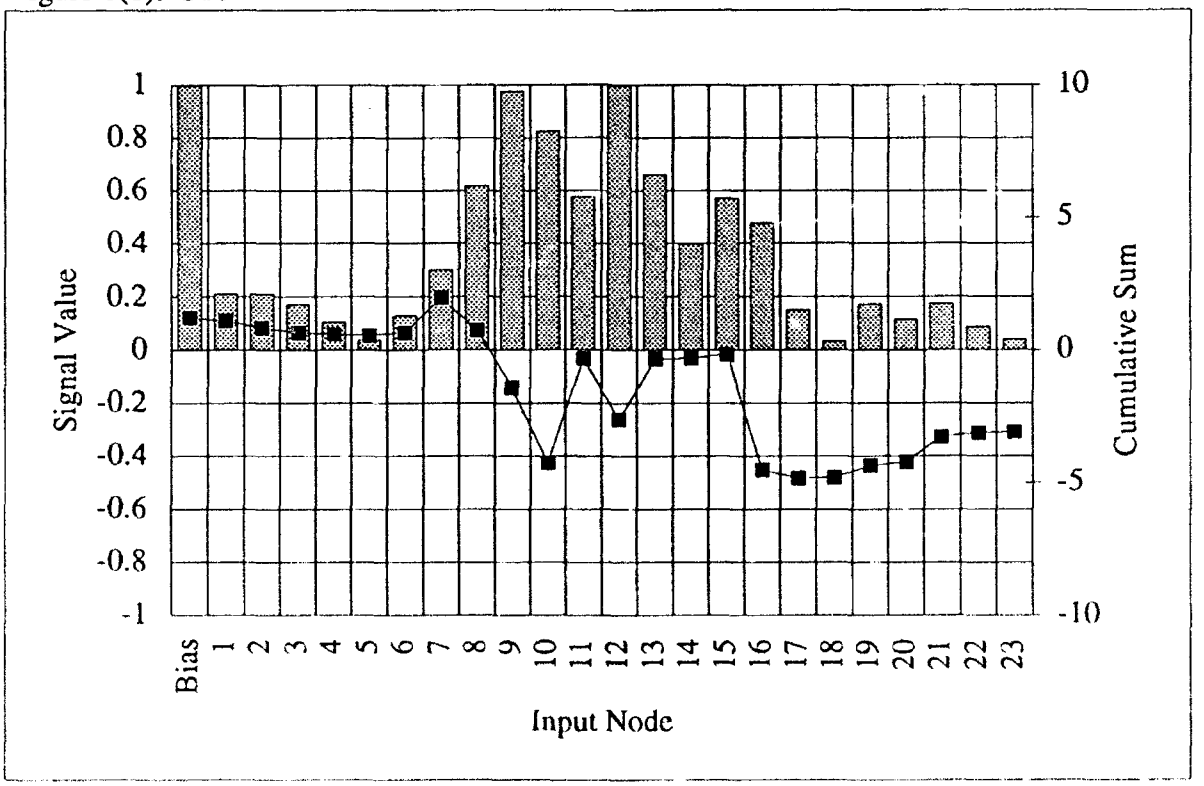


Figure 10.2.1.3-6 Cumulative Sum of Hidden Node Bot4H(1)F-H1
for Instance Nine of Classes B10 and S10

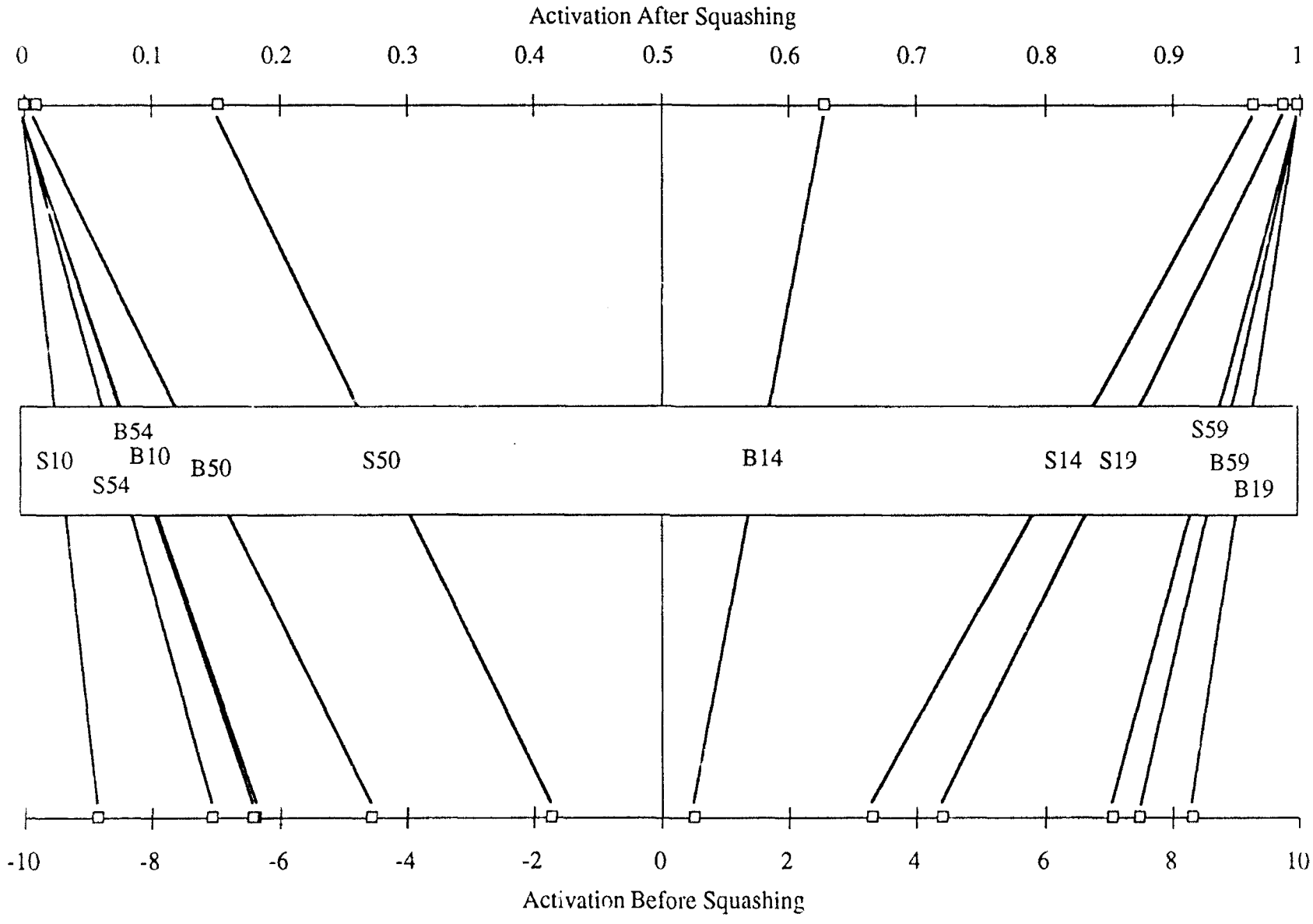


Figure 10.2.1.3-7 Hidden Node Bot4H(1)FN-H1 Activation for Instance Nine of Each Signal Class

node as embedded in the input weights shown in Figure 10.2.1.3-8 is quite simple and quite different than that of F-H1. FN-H1 is sensitive to energy in bins 8 and 9, activating signals with little energy over those frequencies. The results of this strategy depart from the processing of the dimensions in question in producing high activation for the two 45° signals.

10.2.1.4 Discussion of Dimensions and Nodes

T-H3 developed exactly the strategy theorized above for the subjects on these dimensions, that is, reacting to the large return from the 90° signals embedded in the bottom reflection. The subjects found this feature easy to identify, and so did the networks. When the time domain network trained with noise developed a different strategy, the strategy still depended largely on identifying this feature.

In the frequency domain we found network nodes which applied strategies in keeping with at least one of the correlated signal measures, standard deviation. The time domain feature of the 90° signals was reflected in certain frequency domain characteristics, such as the width of the frequency distribution, and the networks were able to extract that information from the signal inputs.

10.2.2 Second Dimensions for Best and N6 Scaling Solutions

These dimensions serve to separate the signals into three groups according to Angle. This is a significant result given the difficulty that subjects had with the angle parameter.

10.2.2.1 Dimensions Analysis

The second dimensions of Best and N6 are almost perfectly correlated (0.99) with one another. A most interesting point about the second dimensions is that the 45° and 0° signals are widely separated. This indicates that, at some level less important than the first dimension, there was a tendency to confuse the signals with other signals of the same angle. Furthermore, the 0° and 45° signals are the most widely separated groups on the second dimensions. To the casual listener this is a surprising result, as the 0° and 45° signals are almost identical. Two subjects, of course, were able to distinguish between them. That performance told us that some features of the signals

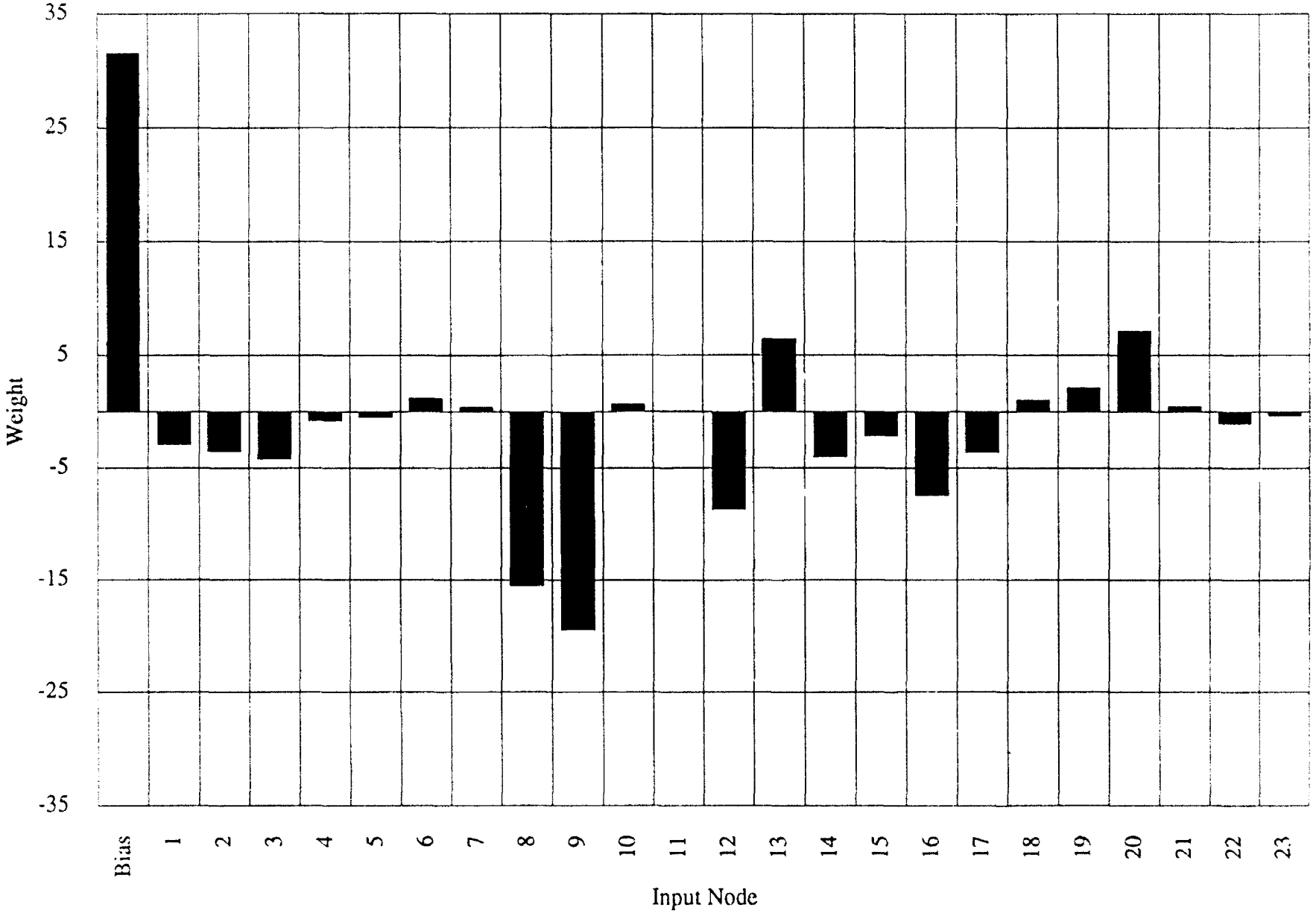


Figure 10.2.1.3-8 Weights on Input Layer to Hidden Node 1 Connections in Bot4H(1)FN

differed. The casual listener might suspect artifact, as no common feature is apparent and only two of ten Navy subjects, and no student subjects, had such high performance.

The breakdown of signals on the second dimension discounts the artifact theory, since the 0° signals group together separately from the 45° signals. The subjects in these scaling runs tended to confuse the 0° subjects with one another, and the 45° subjects with one another, with enough regularity to force the scaling algorithm to place the signals in these groups on the second dimension. Had each signal had some unique artifact, subjects would have confused it with the other Angle class (90° excepted) as often as with its own. Furthermore, such artifact could have been used to identify the signal on other parameters, but those performances remained low.

Listening to the second dimensions was revealing. The discrimination between 0° and 45° signals was quite difficult, as shown by the performances on the experiment. The first-time listener is unable to discern any difference. Armed with the knowledge that two subjects had been able to do the task, two authors sought a feature by which the task could be accomplished. The first author listened to the 0° and 45° signals at 32 kHz, twice the rate at which the signals were played in the experiment. After very considerable time listening to the signals, the author developed a theory about a feature by which the two groups were distinguished. The theory stated that the 45° signals contained an event, similar to the reflection of the 90° signals, but of vanishingly small amplitude. In an informal test the author was able to identify 75% of a test set consisting only of 0 and 45° signals correctly, and the 25% incorrect classifications were on the same two signal in every case.

Upon attempting to apply this theory at 16 kHz, however, the author found that the signal feature was not present. When played at 16 kHz and at the same loudness the signals did not have the same feature. The shift in frequency had uncovered, or made apparent, a signal feature not evident in the signals at the lower frequency. A second author attempted to find the feature and failed. However, that author increased the loudness of the signals (by adjusting the volume of the receiver) and discovered another, probably related, feature. According to this theory, the 45° signals contained two pulses similar to the 90° pulse but of far smaller amplitude, while the 0° signals contained only one. Armed with this description of the features, the first author took the formal test session of the experiment and scored 91 correct Angle classifications out of 96. Such a score indicated that the feature was indeed present in the 0 and 45° signals, and was simple enough to explain. The feature was dependent on loudness level, appearing only when loudness was rather high.

Just as no signal statistic was correlated with either second dimension, no statistic was significant as a predictor in a regression equation. This is not surprising given the subtlety of the signal feature which distinguishes 0° signals from 45° signals. None of the signal statistics would be expected to react to this feature.

10.2.2.2 Hidden Nodes

While some hidden nodes were correlated with these dimensions, the extremely subtle feature which only two Navy subjects found was presumed lost with the information eliminated from the signals in preparation for network input.

10.2.3 Summary

In the Bottom data were found perhaps the closest relationships between human and network processing. On the first dimensions human and networks applied the same strategy to detecting 90° signals, namely, by searching for the large transient characteristic of the broadside orientation. Frequency domain hidden nodes were sensitive to a related feature, demonstrating the network's ability to find signal features to which humans are less sensitive.

While we saw that Navy subjects often performed better than subjects without sonar background, the placement of signals on the second dimension by the subjects who were able to tell 0° from 45° signals is perhaps more impressive. This feature eluded all other subjects as well as the neural networks, and demonstrated a limitation of networks in learning very subtle patterns. Different signal representations might have been an aid to networks in this respect.

10.3 DIMENSIONS OF THE AIR SIGNALS

The two scaling solutions, for three subjects (called Best) and for subject N4 alone, show considerable similarity. The placement of signals on these dimensions was shown in Figures 7.4.3-1 and 7.4.3-2. The relative placement of the signals on the dimensions is quite significant to the analysis of the dimensions. Both first dimensions separate the signals by Thickness. In fact, the first dimension of N4 does so perfectly. The third dimension of N4 divides the signals perfectly by Material, while the third dimension of Best does so with one error. There are,

however, no very good breakdowns of the signals by Striker on any dimension. Every test subject scored higher on the parameters of Thickness and Material than on Striker, and the scaling dimensions reflect this performance. One of these dimensions did produce a partial breakdown by Striker. This was the second dimension of N4, who was the subject with the best performance on Striker. On this dimension the four metal striker signals are lowest, while plastic and wood striker signals are distributed above the metal striker signals. The relationships among the dimensions, the acoustic signal measures, and the network nodes are shown in Figures 10.3-1 and 10.3-2.

All three dimensions in each scaling solution are weighted significantly by the subjects, indicating that the strategy behind each dimension is of some importance. There are some important correlations between dimensions across the two scaling solutions. The first dimension of the Best solution is highly correlated with the first dimension of the N4 solution, indicating that subject N4 used a primary strategy similar to that of the three Best subjects as a whole. The second dimension of the Best solution is highly correlated with the third dimension of the N4 solution. The remaining two dimensions are independent, indicating some difference between the overall strategies used by N4 and the three Best subjects.

10.3.1 Introduction to Air Time Domain Network Nodes

A large number of hidden nodes in the Air time domain networks had interesting correlations with human subject dimensions, signal statistics, and curve fitting parameters. Of particular interest were the networks Air4H(2)T and Air4H(2)TN. The general analysis of these two networks is introduced here in preparation for later sections in which specific hidden nodes are addressed.

These networks have identical architectures, and were trained from the same initial conditions. Their training differed only in that the latter was trained with, the former without, noise added to the signal set. This resulted in the evolution of very different weights in the two networks. Despite the differences, it was frequently the case that a pair of hidden nodes, one from each network, would correlate strongly with the same parameters and with each other. The hidden nodes of interest were thus analyzed in pairs, in order to gain insight into the role played by noise in training.

The output weights of these two networks are shown in Figures 10.3.1-1 and 10.3.1-2. In both, a pseudo-binary relationship developed between the M and W output nodes. When a pseudo-binary

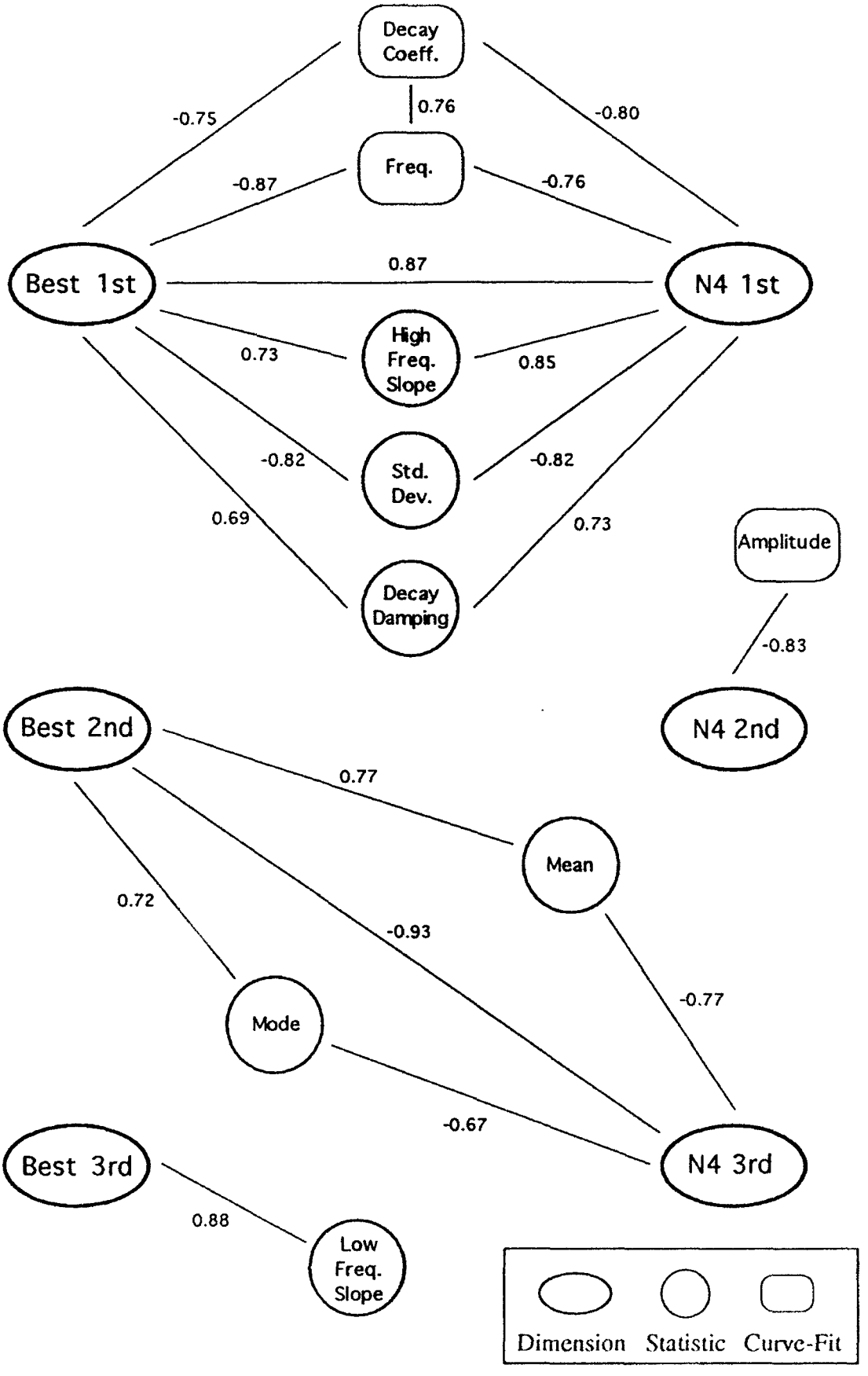


Figure 10.3-1 Correlations Among Air Scaling Dimensions and Signal Measures

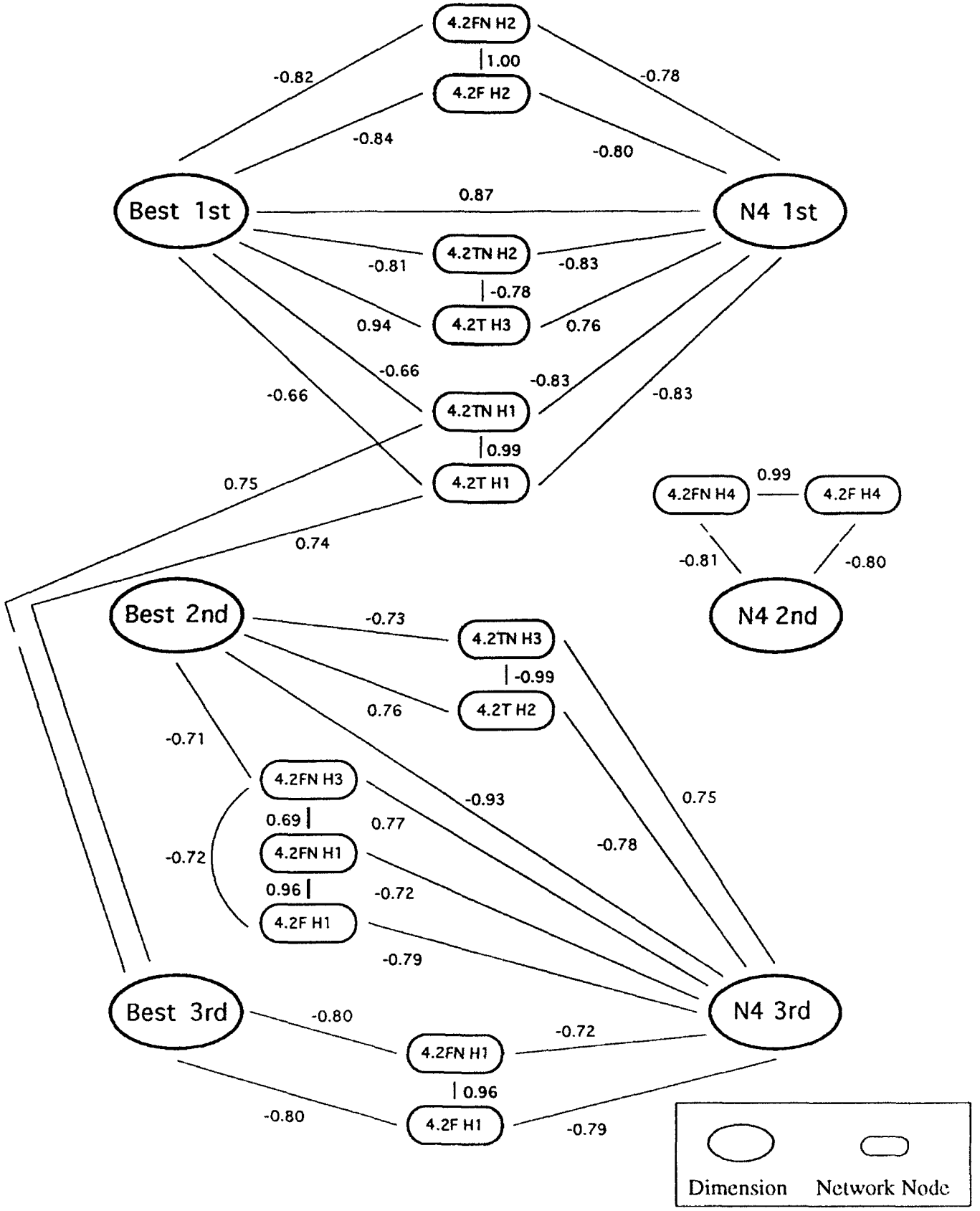


Figure 10.3-2 Correlations Among Air Scaling Dimensions and Network Hidden Nodes

Figure 1(a)

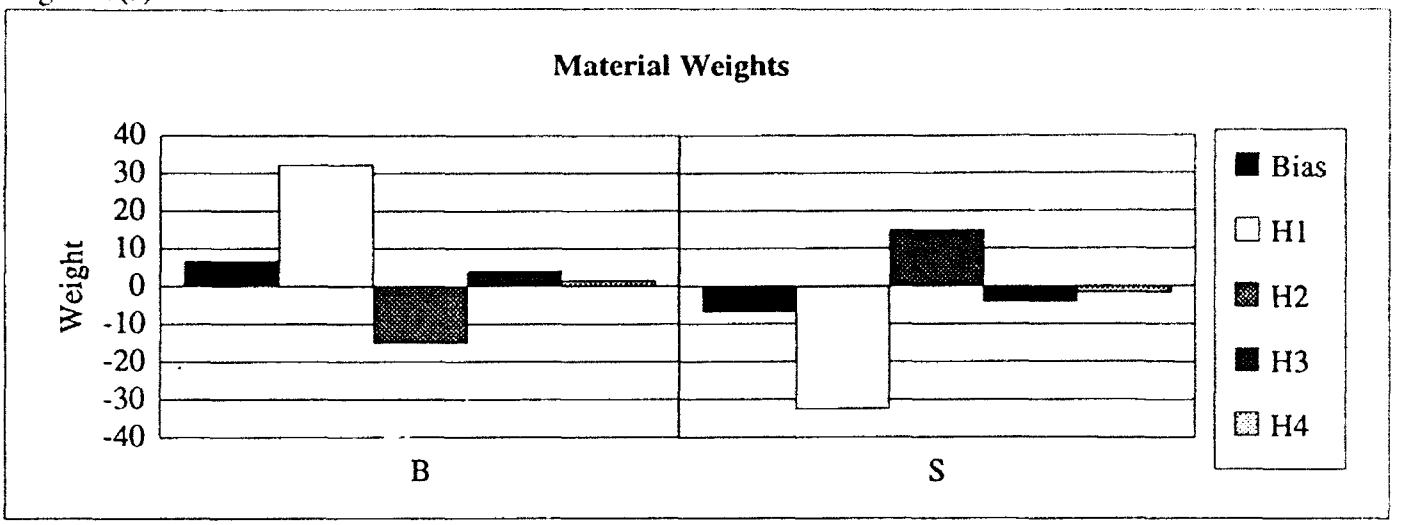


Figure 1(b)

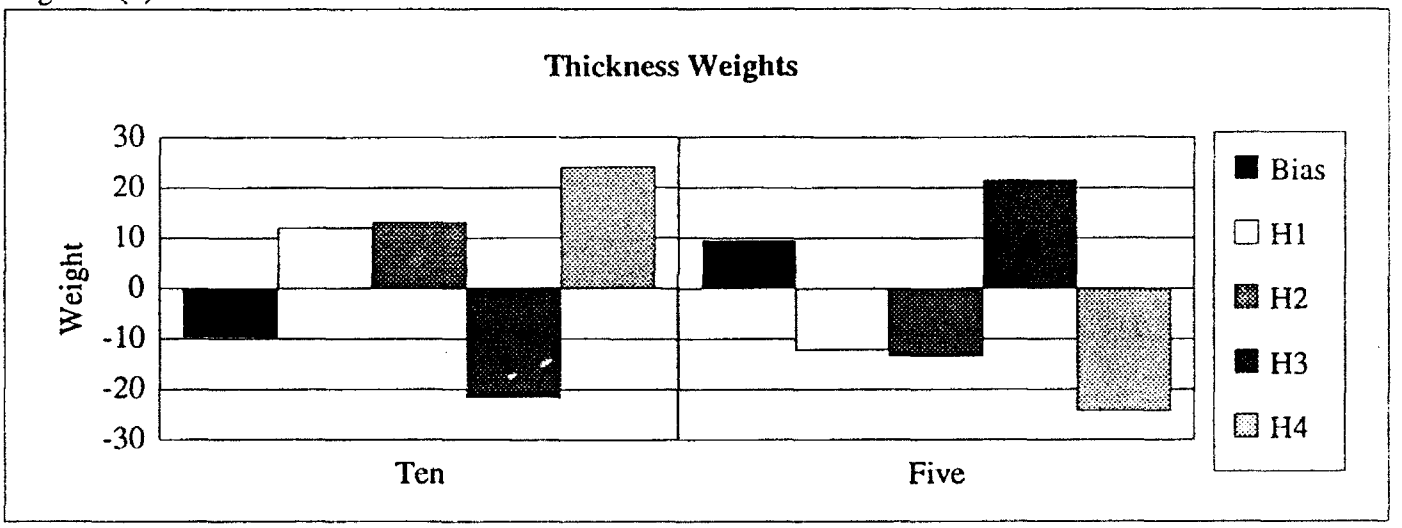


Figure 1(c)

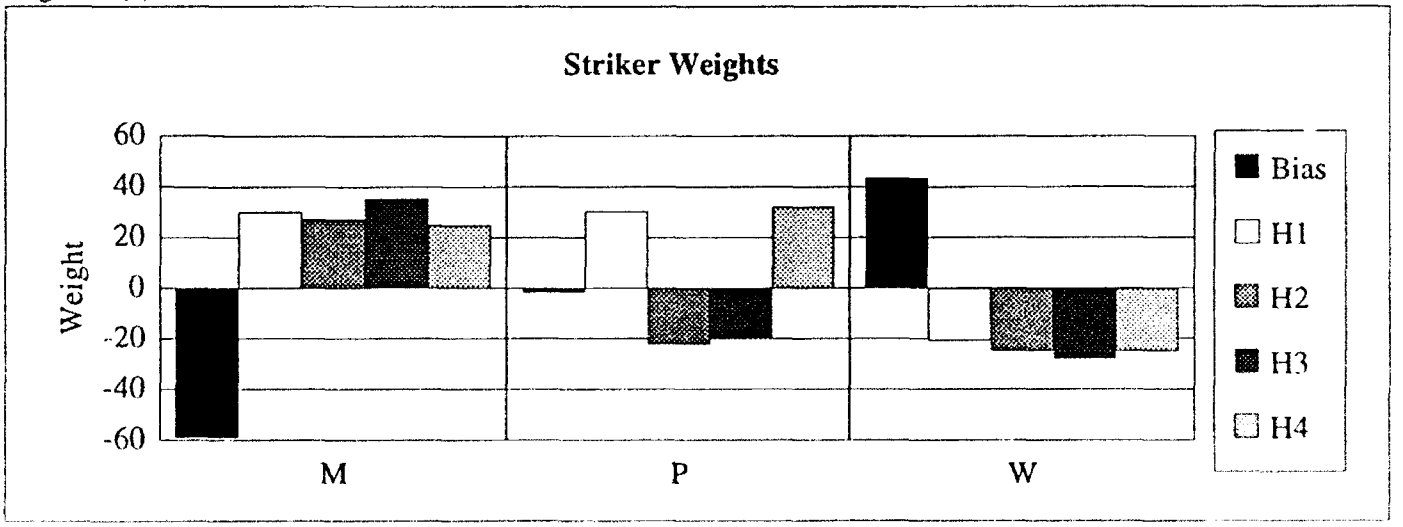


Figure 10.3.1-1 Weights on Hidden to Output Layer Connections in Air4H(2)T

Figure 2(a)

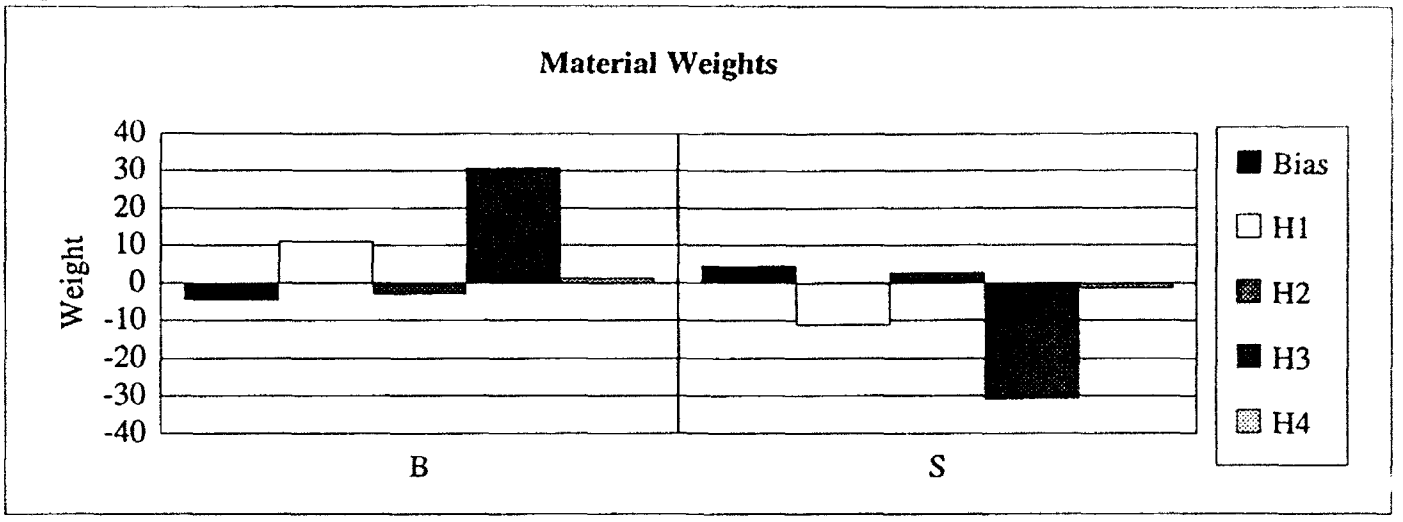


Figure 2(b)

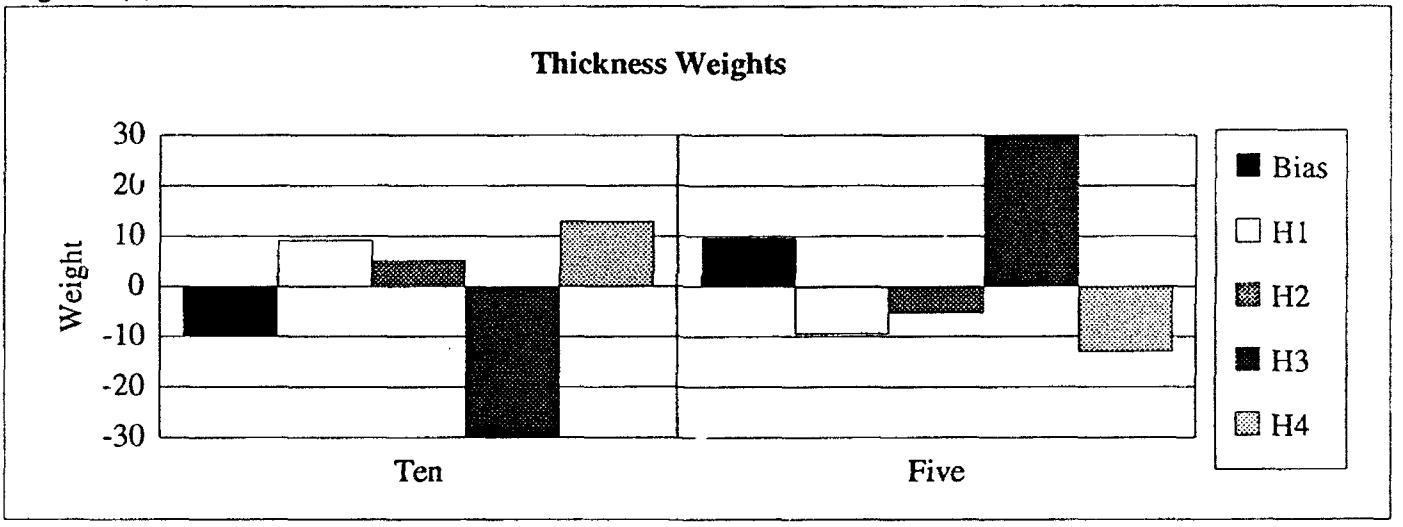


Figure 2(c)

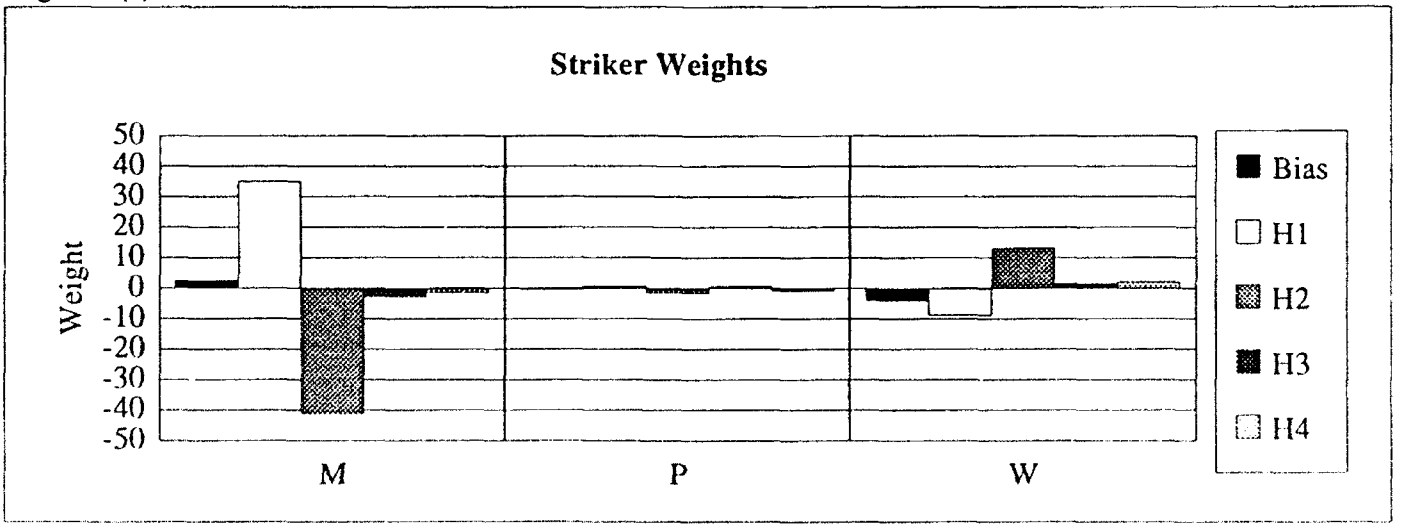


Figure 10.3.1-2 Weights on Hidden to Output Layer Connections in Air4H(2)TN

relationship exists between two of the Striker outputs, the third may not actually perform a meaningful calculation. For example, in Air4H(2)TN (see Figure 10.3.1-2(c)), the hidden node weights of output node P are so small that its activation hovers near 0.5, regardless of the type of signal applied. If the other Striker nodes, M and W, are both suppressed, the network may (correctly or not) place a signal in the Plastic Striker category, but this does not change the fact that the network did not learn actually to identify Plastic Strikers.

On the other hand, in Air4H(2)T (see Figure 10.3.1-1(c)), it is likely that the output node P did learn to identify Plastic Strikers. The weights are distinct from those of the other two Striker nodes, and of respectable magnitude. Despite the pseudo-binary relationship of the other two nodes, the third node here performs a useful function.

During the analysis, the following characterization of the Air time domain signals will be useful. The envelope of each Air signal was observed to conform to one of three qualitatively different shapes. The first, a "short envelope" is one which decays monotonically from its maximum value to very small values within the first twelve inputs. Short enveloped signal types are B1M, B1P, B1W, S5M, and S5P. A "long envelope" signal rings out, having energy at least as far out as input twenty-five. Usually, these signals do not decay according to a single exponential; rather their envelopes may have bumps and plateaus. This group consists of B5M, B5P, B5W, and S1M signals. The third group is characterized by an initial, rapid decay to small values, followed by one or more "returns" of signal energy. The members of this class are S1P, S1W, and S5W, and they are called "boomerang" signals. These categories are introduced only for descriptive purposes, not as a definitive or rigorous categorization scheme.

One minor technicality concerning the Air signal sets should also be commented upon at this point. In the time domain the Air signals all begin with the sharp impact of the Striker, and hence start at their maximum value and decay from there. Since each network input in the time domain was normalized to its maximum value, the first time input in every signal has a value of 1.0, regardless of its signal class. In networks trained with noise, the first input will in general be changed by the noise, but in clean-trained networks, the first input is fixed at 1.0, and hence behaves exactly like a second bias input. When analyzing clean networks, then, any connection weight from input node I1 can for the purposes of analysis be added to the bias weight. The term "effective bias" will be used below to refer to the combined value of the weights on the bias and first input I1.

10.3.2 Best First and N4 First Dimensions

These dimensions were analyzed using a combination of correlated signal measures, regressions, listening to the signals in the order found on the dimensions, and finally network hidden nodes. The first dimensions are considered somewhat, but not overwhelmingly, more important in their respective scaling solutions than the remaining dimensions (based on subject weighting reported in Section 7). Both dimensions separate the signals by Thickness as shown in Figures 7.4.3-1 and 7.4.3-2. The Best first dimension does so with two errors near the center of the dimension. The N4 first dimension separates the 10% signals perfectly from the 5%, although two 10% signals are placed very close to the 5% group rather than with the remaining four 10% signals. The three Brass 10% signals and S1W are together low on the dimension, while S1P and S1M are very close to the group of 5% signals high on the dimension. The Best first dimension differs in that S1P is part of the 10% group low on the dimension, and S5W is in the middle of the dimension rather than high. The Brass 10% signals are low on both first dimensions, suggesting that these signals share some feature to which all three subjects were sensitive and which differentiates them from the bulk of the rest of the signals.

10.3.2.1 Dimensions Analysis

The two first human scaling dimensions are highly correlated with several statistical measures, as seen in Figure 10.3-1. Statistics taken in both the time and frequency domains correlate with these dimensions. Among the curve fit parameters, both the decay coefficient and the frequency of the most persistent sine wave (i.e. for each signal, the sine wave which damps at the slowest rate) are highly negatively correlated with the dimensions. This indicates that as the value of the signal on the dimension increases, the most persistent sine wave of that signal tends to last longer than that of other signals, and tends to be of lower frequency. The high correlation with the time domain decay damping statistic is consistent with the correlation with the damping coefficient of the curve fit solution. High frequency slope and standard deviation, two statistics which characterize the shape of the frequency distribution, are also correlated with the first dimensions. The correlation with high frequency slope indicates a sharper cutoff of high frequencies for signals higher on the dimension. Signals high on the dimension would also appear to have a wider distribution of frequencies than signals low on the dimension.

Upon listening to the signals according to their placement on the first dimension of the N4 scaling solution, the first impression on the listener was a time domain difference between the two groups of signals. The large group of signals high on the dimension damp much more slowly than the signals low on the dimension. The Brass 10% signals grouped low on the dimension are quite distinct in damping faster than all others. S1W is an exception to this rule. S1W is unique in having both a distinct, dull strike and a long ring. If the placement of S1W near the Brass 10% signals was due to its distinct, dull strike, as seems feasible, then the subject was listening for decay only from the initial frequencies of the strike. In these ways the high correlations with the decay coefficient of the most persistent sine wave and with the "decay damping" statistic are apparent to the listener.

Listening to this dimension is also an aid to understanding the negative correlation with the frequency of the most persistent sine wave used in the curve fit solution. As we progress from signals with high dimension values to signals with lower values, the frequency of the long-duration ringing portion of the signal was heard to increase. The exception, again, is S1W, which has a ringing frequency similar to the other Steel 10% signals which are higher on the dimension. The effect is not linear with the frequency in Hz, but the nonlinear nature of human hearing along with the complexities of subject strategies would not be expected to give a linear relationship. The order effect is quite good, in that one can hear the frequency differences consistently from signal to signal along the dimension.

The relationship between the first Best dimension and the rates of decay of the signals is also apparent from listening. The signals that damp the fastest are lowest on this dimension, and the relationship is audible. The high correlations with the two damping parameters make sense to the listener. The high correlation with standard deviation also becomes apparent with listening to this dimension. The longer signals are dominated by their ringing portion, which contains far fewer frequencies than the relatively broad spectrum of the impact. Subjects are using some combination of these time and spectral characteristics, which tend to vary together on this dimension. That is, the signals which damp the fastest have the widest frequency distributions, as measured by the standard deviation, precisely because they damp faster than other signals.

Another audible characteristic of the first dimension of the Best solution is that the frequency of the ringing portion of the signals tends to increase as the value on the dimension decreases. The

signals which damp very quickly are more difficult to interpret in this manner since it is hard to identify their longest-lasting frequency, yet they contribute to the correlation quite well.

The best single regression predictor for the first dimension of the N4 solution is the high frequency slope of the signal:

R^2(adj) = 69.1%
High Frequency Slope p = 0.0005

However, this performance is due to the wide separation of the Brass 10% signals from the other signals, which serves to predict only to which of these groups a signal belongs. This is shown in Figure 10.3.2.1-1.

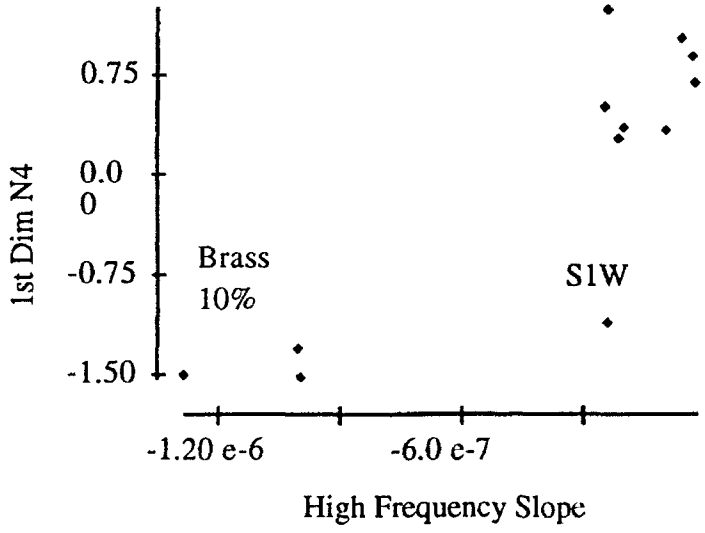


Figure 10.3.2.1-1 High Frequency Slope vs. First Dimension

The decay coefficient used as a predictor separated the two groups of signals in much the same manner. A more revealing regression model was created from the frequency of the most persistent sine wave. This predictor was not as strong statistically as high frequency slope, but had a better distribution of the signals, as seen in Figure 10.3.2.1-2.

R^2(adj) = 52.7%
Frequency p = 0.0045

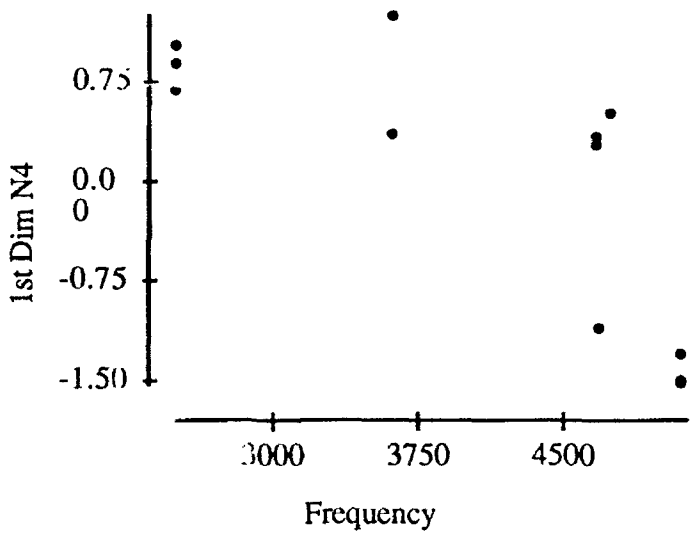


Figure 10.3.2.1-2 Frequency of Most Persistent Sine Wave vs. First Dimension

Here we see a relationship between the dimension and a predictor which spans the range of the dimension. The signals are no longer simply clumped in widely-separated groups. Of course this does not account for non-linearities in the subjects' perception of frequency or in their placement of signals on the dimension by frequency, but offers an explanation for the placement of intermediate signals on this dimension not offered by the previous regression predictors.

The three subjects who made up the Best group may have been using frequency in a more straightforward manner, as this parameter of the signals is a better predictor than it was for N4:

R^2(adj) = 73.7%
Frequency p = 0.0002

Again, subject perceptions of frequency are not fully accounted for by such a simple model, and the relationship does not appear to be linear, but this statistic is a very good predictor of placement on the first dimension of the Best solution.

10.3.2.2 Analysis of Air4H(2)T-H3 and Air4H(2)TN-H2

The first set of nodes to be discussed are both from the Air signal, four hidden node, time domain networks. The second and third hidden nodes for these networks are referred to by the names Air4H(2)TN-H2, and Air4H(2)T-H3. This pair of hidden nodes was chosen because of very significant correlations between each of them and the first dimensions produced in the human performance analysis. They are also correlated with each other, yet they have very different weight structures and so respond similarly to the signals through rather different means. Since the following discussion applies only to these two networks, to simplify the notation they may be further abbreviated from Air4H(2)T and Air4H(2)TN to simply T and TN. Nodes within the networks will be referred to in a similar manner, for example T-H3.

Following the procedure outlined in Section 10.1.1, the analysis will begin at the output layer. Figures 10.3.1-1 and 10.3.1-2 show the hidden-to-output weights of the networks T and TN, respectively. A comparison of the two reveals that the major difference in the output layer between the two networks occurs in the Striker weight structure. There are significant quantitative differences in the Material and Thickness weights as well, but only in the Striker weights have the networks developed qualitatively different weight structures. Focusing on the nodes of interest, T-H3 and TN-H2, the weights connecting these two nodes to their respective output layers follow nearly opposite trends. For example, TN-H2 has a strong negative connection to TN–M while T-H3 has a strong positive connection to T–M. This is not surprising since the negative correlation between T-H3 and TN-H2 implies that they tend to sort the signals into opposite orders. Prior to any further comparison of nodes T-H3 and TN-H2, it will be useful to continue the analysis of each node individually.

10.3.2.2.1 TN-H2 Analysis

First, consider hidden node TN-H2 and the weights connecting it to the output layer (see Figure 10.3.1-2). The Brass output node TN–B weights TN-H2 negatively, but very weakly compared to its weights on TN-H1 and TN-H3. In fact, TN–H2 receives a weight smaller even than the bias term. From this it may be inferred that TN–H2 is not a primary node used for determining target material. The situation is similar for TN–Ten; it places a positive weight on TN–H2 which is small compared to all the other hidden node weights and the bias term. Thus, it would seem that the Thickness outputs are also largely unresponsive to TN–H2.

Turning now to the Striker output nodes, it is evident that TN–H2 plays a key role in the determination of the striker. The weights developed by TN–M and TN–W display the pseudo–binary tendencies described in the introduction to this section. The same hidden node values which produce activations in TN–M will tend to suppress TN–W, due to the opposite and roughly proportional weights these nodes place on the hidden layer. For all three Striker output nodes, the weight placed on TN–H2 is larger in magnitude than the weights from any of the other hidden nodes. In particular, TN–M and TN–W place upon it an extremely high weight, negative and positive respectively. Looking at the other weights between the Striker outputs and the hidden layer suggests that TN–H1 also plays a role in determining the Striker. A precise understanding of how the Striker is determined would involve at least these two nodes. For the present discussion of TN–H2, however, it suffices merely to know that it is heavily used by the network as a Wood detector and a Metal rejector, and is not used much by other output nodes.

The weights connecting TN–H2 to the input layer are shown in Figure 10.3.2.2.1-1. There are two features of this weight structure which simplify its analysis. First, the only important weights connecting TN–H2 to the input layer are concentrated between the input nodes TN–I2 and TN–I8. Outside this range, not only are the weights smaller in magnitude, but the inputs by which they are multiplied are very small, even in long enveloped signals. Second, these weights are uniformly negative, in contrast to the bias, which is approximately equal in magnitude to the largest input weight (TN–I2), but positive. This bias term gives TN–H2 a high activation which is decreased by signal energy in TN–I2 through TN–I8. Only a signal which has sufficient energy in this region, and/or decays sufficiently slowly, can overcome the bias, and turn off TN–H2. Therefore, based on the input weight structure, it may tentatively be concluded that the hidden node TN–H2 is a detector of fast decaying signals.

The above observations of the output and input weights suggest the following description of this node's function. The hidden node TN–H2 provides some information about the striker, using information found early in the signal, with not much regard for the signals' behavior after the first few time inputs. This is rather appealing from a physical point of view; one would expect the impact of the striker to influence most strongly the transient, i.e. quickly damping, components associated with the production of the sound. The "after–ring" is more characteristic of the natural resonances of the target than the striker. Thus it is consistent for TN–H2, which the network uses for classification of the Striker, to focus on the early portion of the signals.

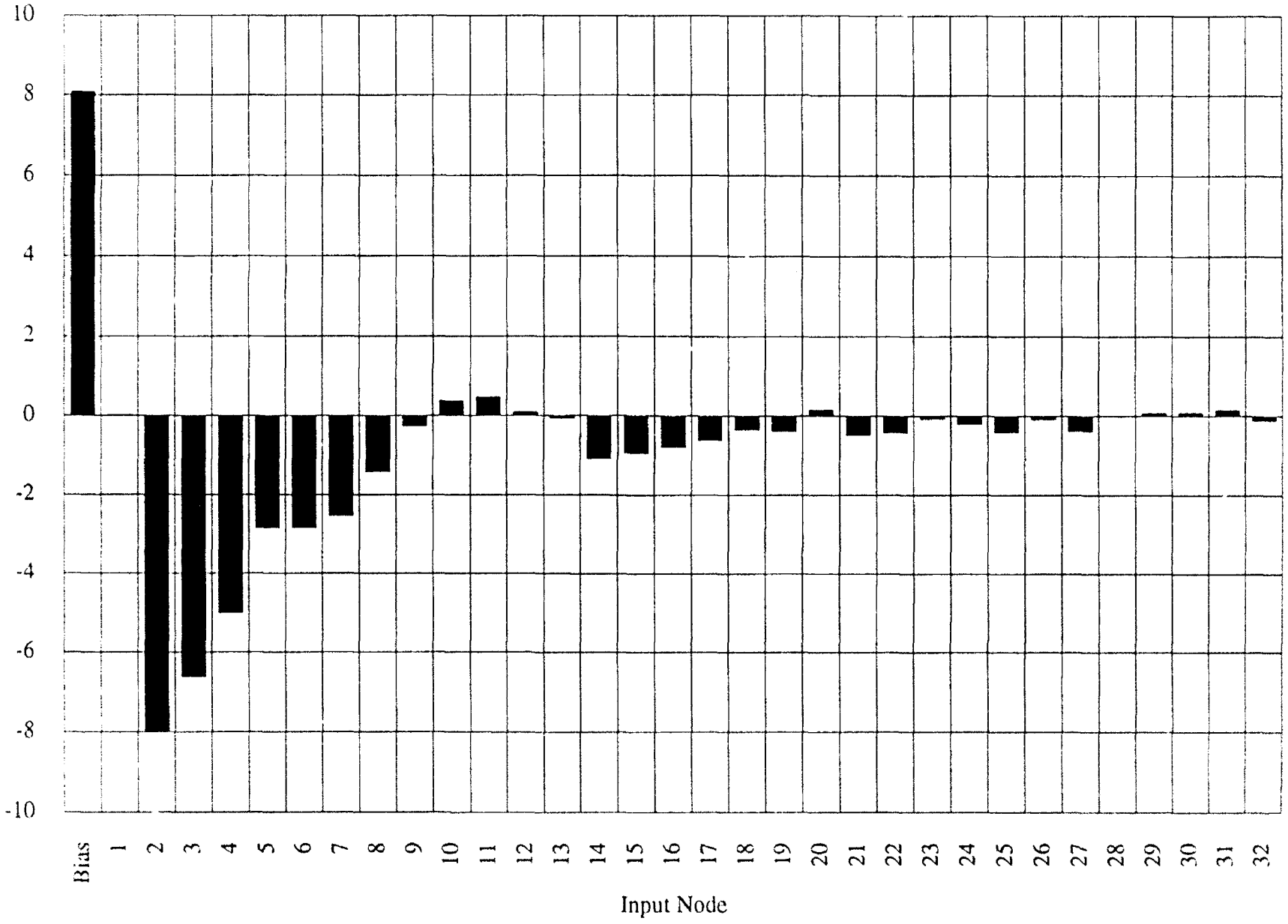


Figure 10.3.2.2.1-1 Weights on Input Layer to Hidden Node 2 Connections in Air4H(2)TN

It was hypothesized above that TN–H2 is a detector of fast decaying signals. By investigating TN–H2's response to actual signals in the Air test set, the validity of this hypothesis can be tested. An output summary is shown in Figure 10.3.2.2.1-2 which displays the activation of TN–H2 resulting from the input of instance nine of each of the twelve signal classes, both before and after applying the transfer function. Plotting the signals in this fashion shows explicitly the effect of the transfer function on the output. It is clear from the result that TN–H2 does not sort the signals perfectly according to Material, Thickness, or Striker. The various Brass signal classes are split, half activating the node strongly and half suppressing it. Because of this it is not at all useful for determining material. Brass signals are separated very well according to Thickness, but different Strikers are clustered together, while for Steel signals the reverse is true. For Thickness, some overall separation of the signals persists, but as remarked earlier, the Thickness output nodes ignore TN–H2. For classifying Striker, TN–H2 fares a little better; three out of four Metal and Plastic striker signals result in negligible activation, while three out of four Wood signals activate this node to some degree. The signals which do not follow this pattern are B1M and B1P, which strongly activate the node, and B5W, which strongly suppresses it. Although the node TN–H2 is only a 75% accurate detector (rejector) of Wood (Metal) signals, TN–H2 is the hidden node most heavily weighted by the Striker output nodes. This may explain why the percentage of correct Striker classifications for the network TN as a whole is only 71%.

To determine what the hidden node TN–H2 has learned about the signal set, it is useful to examine more closely how the node output evolves under the influence of the various network inputs. This is readily accomplished graphically, and since this graphical method will be applied extensively throughout the hidden node analyses, some explanation of the meaning of the graphs will now be given.

The graphs used to view the response of specific hidden nodes to specific signal classes plot two different quantities as a function of input node. One is shown as a column plot, and is simply the value of the signal being applied. The second, shown as a curve, is the cumulative sum of the hidden node. The contribution to a hidden node's cumulative sum made by each input node is the product of that input node's value and the weight connecting that input node to the hidden node. The cumulative sum plotted for a specific input node is the sum of contributions from all the inputs from the bias up to and including that node. The influence of a particular input can be read from the difference in the cumulative sum between that input and the previous input. The graph thus serves to convey how important each successive input is to the final value of the cumulative sum.

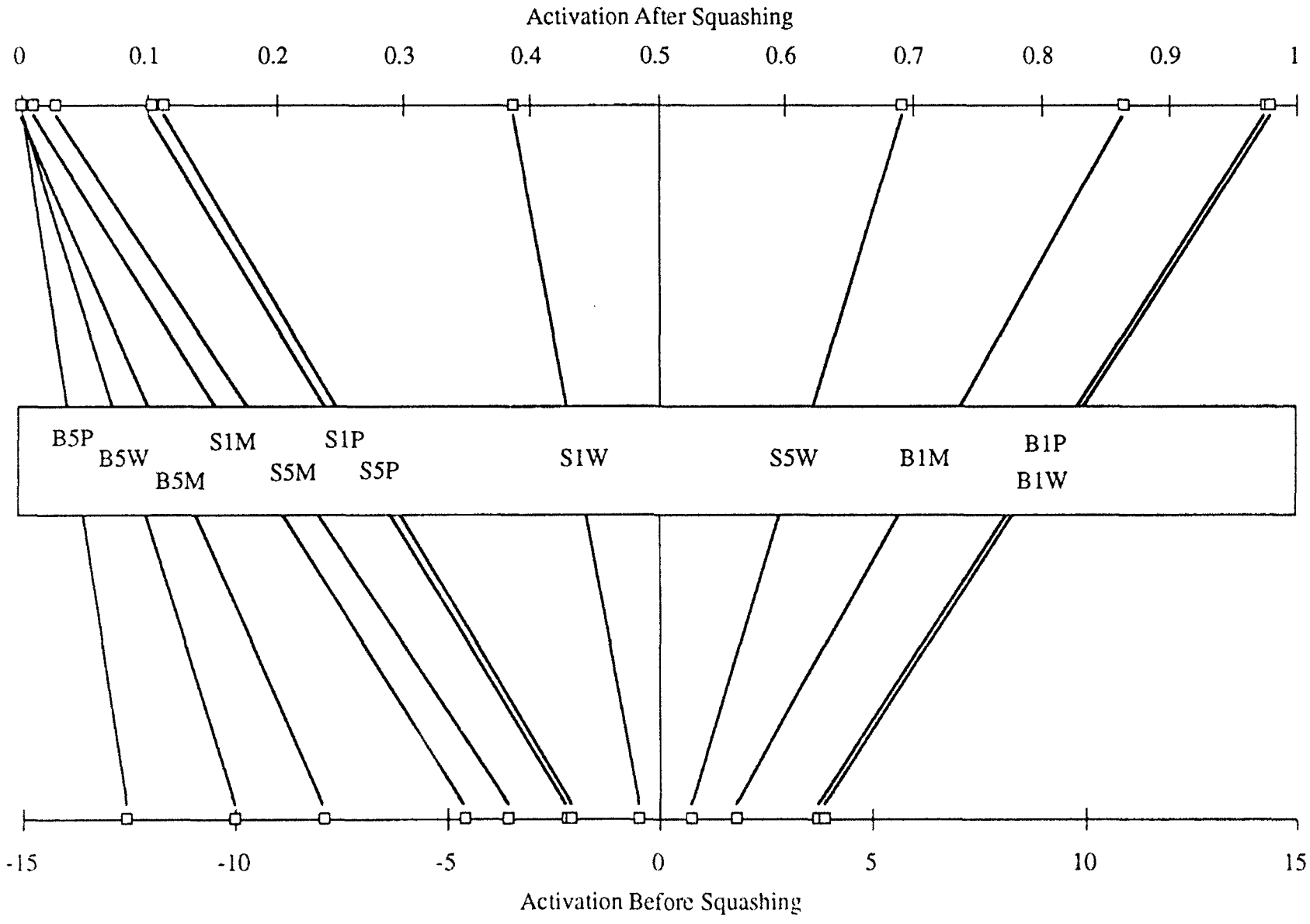


Figure 10.3.2.2.1-2 Hidden Node Air4H(2)TN-H2 Activation for Instance Nine of Each Class

The value of the cumulative sum computed at the last input node is the argument to the transfer function which produces the final output of the hidden node. For example, Figure 10.3.2.2.1-3(a) shows the behavior of the cumulative sum of hidden node TN–H2 when instance nine of the B1P signal class is applied. The first point of the cumulative sum plotted corresponds to the bias, and is approximately +8, this being the product of the bias value of +1.0 (also shown in Figure 10.3.2.2.1-3(a)) and its weight (shown in Figure 10.3.2.2.1-1). Although the next input, TN–I1, is +1.0, the cumulative sum does not change at input TN–I1, because the weight TN–H2 places on this input is 0.0. The drop in the cumulative sum between TN–I2 and TN–I6 is caused by substantial energy present in these negatively weighted inputs. No significant change occurs after TN–I6, due to a combination of small weight values, and low (mostly zero) inputs in this region. The largest single drop in the cumulative sum is approximately 2.5, and occurs at input TN–I2, whose value is approximately 0.3, and whose weight is about –8. The final value of the cumulative sum is approximately +3.8, which corresponds to an output of about +0.98 after the transfer function is evaluated. These are the values shown for this signal class (B1P) in the activation summary shown in Figure 10.3.2.2.1-2.

To return to the analysis of TN–H2, the idea that this hidden node is a detector of fast decaying signals certainly holds true for the most extreme examples in the Air signal set. Figure 10.3.2.2.1-3 shows the cumulative sums for instance nine of the B1P and B5P signal classes, which are representative of the shortest and longest signals, respectively. It is clear from Figure 10.3.2.2.1-3(a) that the B1P signal simply lacks enough signal energy to overcome the bias term, and thus fails to deactivate TN–H2. By contrast, the B5P signal shown in Figure 10.3.2.2.1-3(b) has more than enough energy to overcome the bias, and suppress the node. So it is easy to see why for the longest signals (B5M, B5P, B5W), the output is nearly 0.0, while for the shortest (B1P, B1W), the output is nearly 1.0.

Consider, however, signals from the class B1M. Although they are as short as the other Brass 10% signals, they give a noticeably lower activation of 0.87. More strikingly, the other short enveloped signals (S5M and S5P) actually have enough extra energy to suppress TN–H2. Thus, while TN–H2 does seem to detect fast decaying signals, only the very shortest signals manage to be detected. This may indicate that it is performing some additional, more subtle function as well. Some insight into this additional operation may be gained by examining the remaining signal classes, which all give moderate activations: S5W, S1W, S1P and S5P. It is noteworthy that this

Figure 3(a): B1P

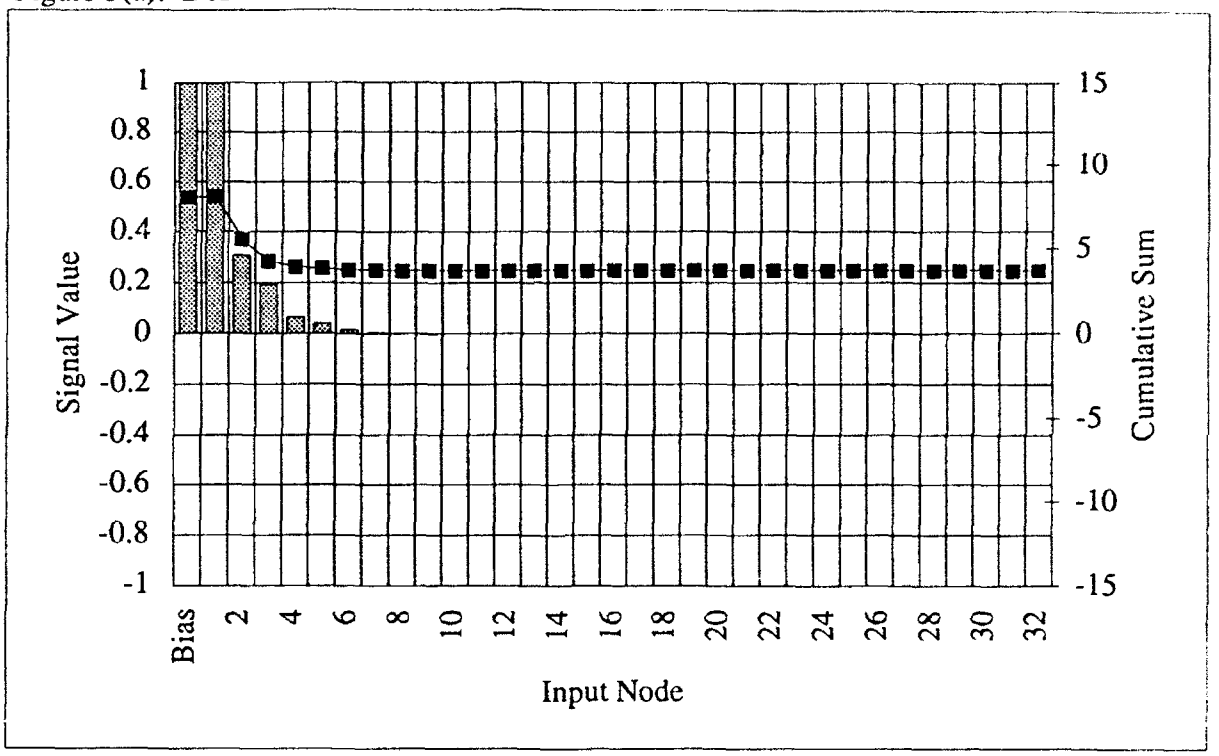


Figure 3(b): B5P

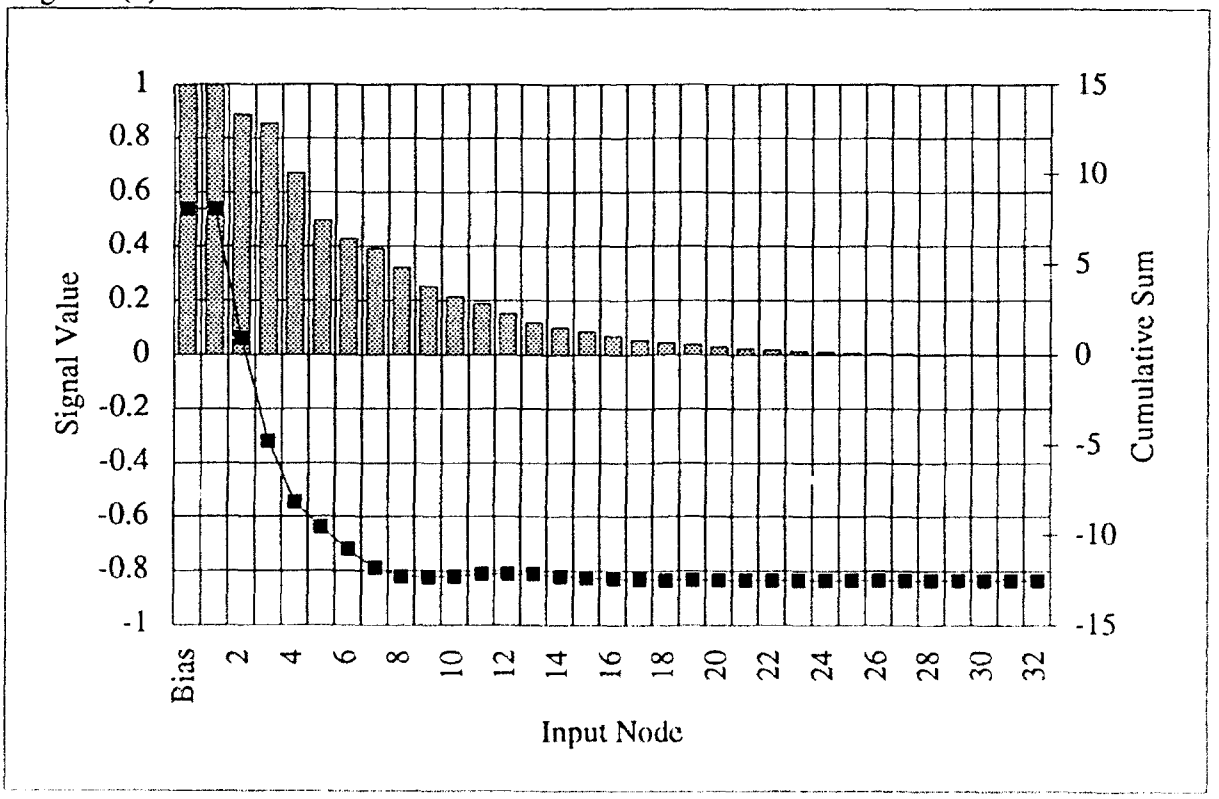


Figure 10.3.2.2.1-3 Cumulative Sum of Hidden Node Air4H(2)TN-H2 for Instance Nine of Classes B1P and B5P

group includes all the boomerang signals. Of this last group, the highest activations occur for the S1W and S5W signals, shown in Figure 10.3.2.2.1-4. The first three inputs of S1W signals show a very strong decay (see Figure 10.3.2.2.1-4(a)). If the signal continued to drop monotonically after TN–I3, its cumulative sum would level off at a high activation level, like the B1P signal shown in Figure 10.3.2.2.1-3(a). Instead, however, in inputs TN–I4 through TN–I13, the first "return" of the boomerang contained enough energy to suppress the node weakly. The return of signal energy in the S5W pattern in Figure 10.3.2.2.1-4(b) is somewhat weaker, but still enough to give it a noticeably lower activation than the B1M signals it resembles for the first few inputs.

In conclusion, then, it may be said that this node is sensitive to a physical quality of the signals, namely, the speed of their decay. It is strongly activated for very short signals, suppressed by long ones, and signals between these extremes are placed in the middle. Long signals are produced predominantly by Plastic and Metal strikers, hence the output layer uses TN–H2 as a detector of Wood striker signals. The hidden node TN–H2 does not perform this function perfectly, which is probably partly responsible for this network's mediocre success with classifying Striker.

10.3.2.2.2 T-H3 Analysis

The performance of the cousin of this network, T, which was trained on clean signals is somewhat better, in that it achieves a level of 84% correct for the striker parameter. The node T–H3 has a significant negative correlation with TN–H2, which suggests that these two hidden nodes sort the signals into roughly opposite orders. It might therefore be expected that this hidden node would be used for similar tasks, but in an opposite manner to the hidden node TN–H2 discussed above. This is true to a point, but there are some major differences between the two networks in the structure of their output weights.

Moving now to the weights connecting T–H3 to the output layer (see Figure 10.3.1-1), the Brass output node gives positive weight to T–H3, but it is much smaller than the bias term. Thus T–H3, like TN–H2, does not seem to be a very important node in determining Material. There is a large, positive weight connecting T–H3 to the Five Percent output, T–Five, which would suggest that the hidden node is used partly as a Five Percent Thickness detector. This is in contrast to hidden node TN–H2, which was ignored by the Thickness nodes. The output nodes T–M and T–W display an

Figure 4(a): S1W

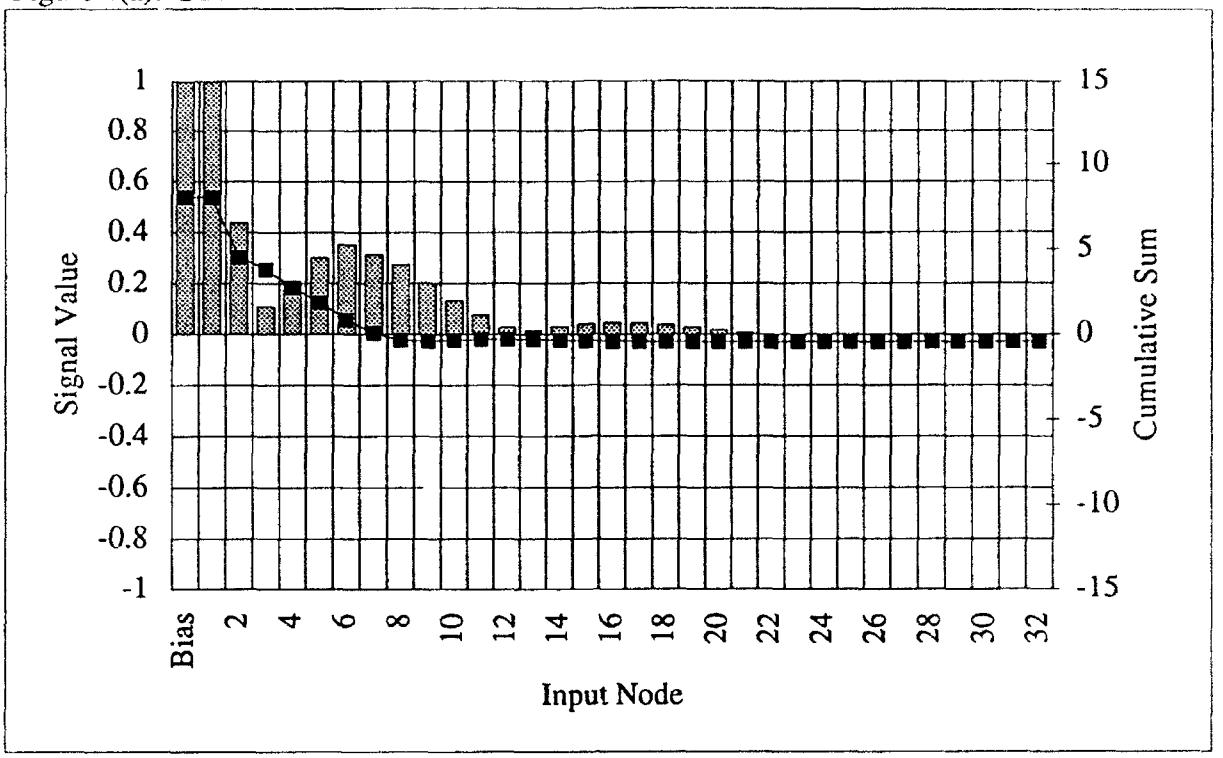


Figure 4(b): S5W

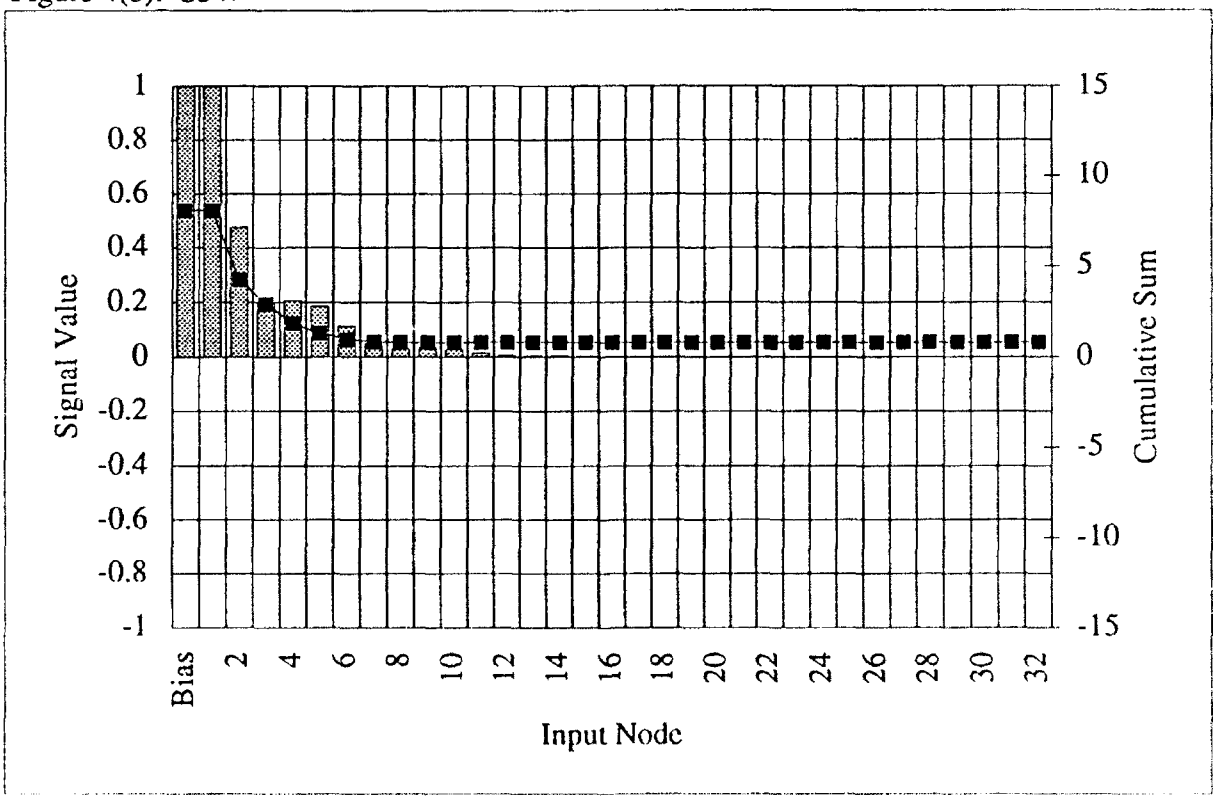


Figure 10.3.2.2.1-4 Cumulative Sum of Hidden Node Air4H(2)TN-H2 for Instance Nine of Classes S1W and S5W

even stronger pseudo–binary relationship than TN–M and TN–W. In determining Striker, the positive weight connecting T–H3 to T–M is larger than all others, save the bias. There are significant negative weights connecting T–H3 to T–P and T–W (the Plastic and Wood nodes). Thus T–H3 is used as a Metal detector by the network. This much is similar and opposite to the usage of TN–H2, which detected Wood signals, and strongly rejected Metal. A difference between the two hidden nodes is that T–H3 is also used by the Plastic output node, T–P. It should be recalled that in the network TN, the Plastic output node did not develop a meaningful algorithm. The additional uses of T–H3 are the most significant differences between the hidden nodes TN–H2 and T–H3.

We now continue to the connections between T–H3 and the input layer of the network T. These weights are shown in Figure 10.3.2.2.2-1. A superficial comparison of this graph and Figure 10.3.2.2.1-1 suggests that the two nodes extract very different features from the signals. Further comparison will be deferred for the moment, however, so that T–H3 can be discussed on its own merits. Since this is a clean-trained network, the bias and first time input may be added (see comments in the introduction in Section 10.3.1) to give an effective bias of approximately –8.0; the node thus starts out deactivated. After T–I1, the weights fall naturally into three groups. The first consists of a complex alternating weight pattern from T–I2 through T–I7. Next follows a simpler group of negative weights from T–I8 through T–I13. The third group consists of the all positive weights from T–I14 through T–I30.

The last group (T–I14 through T–I30) is the easiest to understand. In all but the longest signals, the inputs to this group are all 0.0. Although these weights are substantial, the longest signals in this region are of small amplitude, hence the contribution from this group is significant and positive, but not overwhelming. This last group can be thought of as a moderately strong long signal detector.

There are not as many weights in the middle group (T–I8 through T–I13) as in the last, but they are larger in magnitude. In addition, the signal in the middle region is much larger than in the last region. The negative contribution from this group tends to overshadow the positive contribution from the last group, and can be considered a very strong rejector of medium or long signals.

The first group (T–I2 through T–I7) is used to process the most energetic portion of the signal, and is very important in determining the final state of the node, but it is also the most difficult to

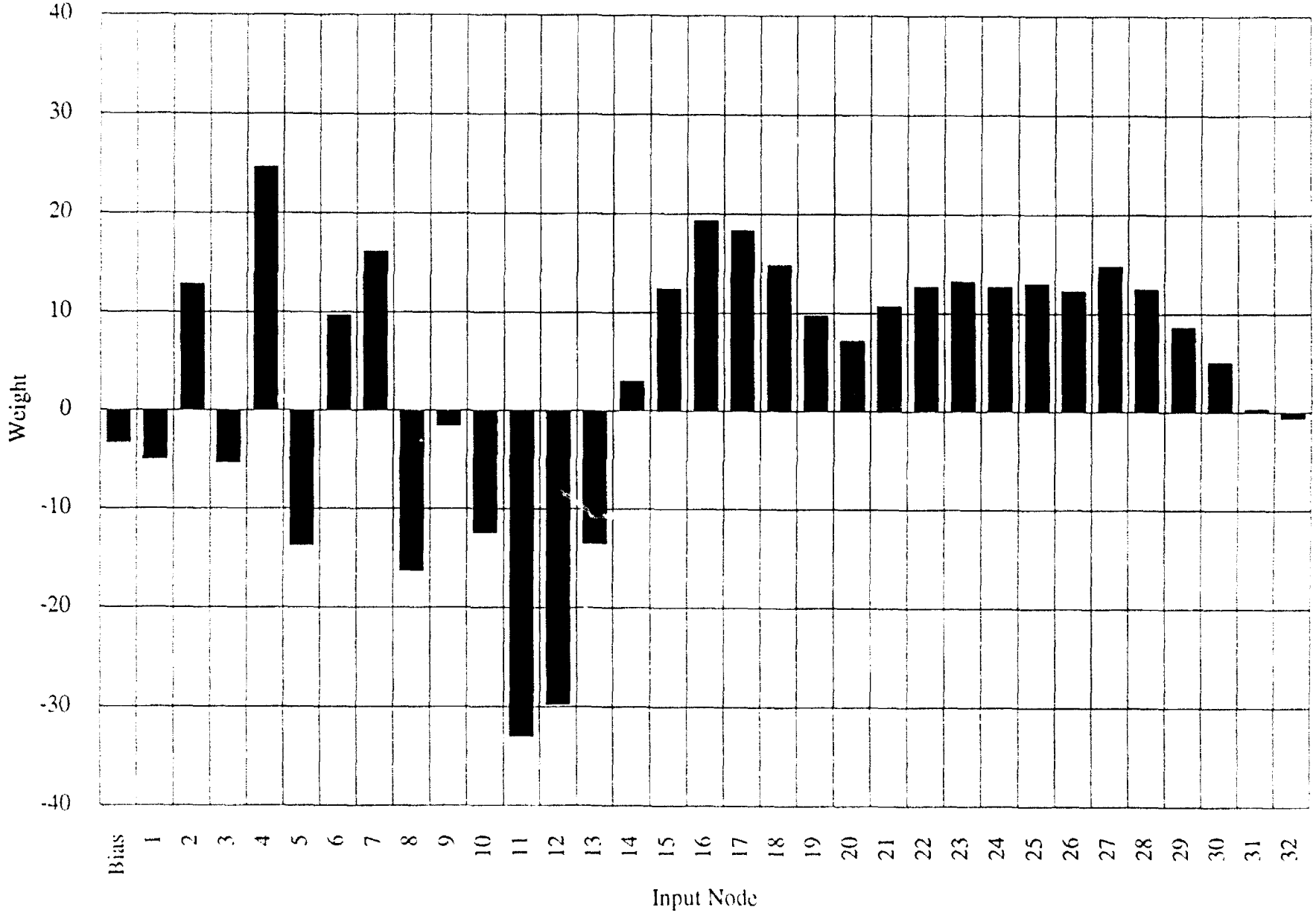


Figure 10.3.2.2.2-1 Weights on Input Layer to Hidden Node 3 Connections in Air4H(2)T

understand. Some simplification results by mentally grouping the weights in pairs: T–I2 with T–I3, T–I4 with T–I5, and T–I6 with T–I7. As shown in Figure 10.3.2.2.2-1, the positive weight on input T–I2 is of significantly larger magnitude than the negative weight on T–I3. Similarly, T–I4 is much more heavily weighted than T–I5. The weights on T–I6 and T–I7 are both positive. This disposition towards positive weights is such that each pair yielded a net positive contribution to the cumulative sum, for all signals applied. This contribution was largest for signals with consistent energy throughout these inputs, and smallest for signals with low energy. It is interesting that both negative weights correspond to the positions of minima, T–I3 and T–I5, in boomerang signals. This seems to be more than accidental, for it helps boomerang signals to achieve higher cumulative sums than short enveloped signals in this region. This first group of weights thus seems to sort signals into long enveloped (highest cumulative sum), boomerang (smaller cumulative sum) and short enveloped (smallest cumulative sum) signals. This group performs a very similar function to that performed by TN–H2. In fact, the cumulative sums obtained from the signals using only this first group of weights (ignoring inputs T–I8 through T–I32) are distributed in almost exactly the opposite order as the sums using all the weights in hidden node TN–H2.

The activations of T–H3 after applying instance nine of each of the signal classes to the input layer are shown in Figure 10.3.2.2.2-2. The placement of the signals is mostly consistent with the negative correlation between this node and TN–H2. The (mostly) subtle differences cause these two hidden nodes to have markedly different functions in the networks. With the exception of S1M and S5W signals, T–H3 separates signals very well according to target thickness, as anticipated from the fact that it is used as a 5% detector by the output layer. There is no separation between Brass and Steel signals, however. For classifying Striker, the node seems slightly worse than its counterpart, TN–H2. It is odd that T–P weights this node heavily, since half of the Plastic signals activate the node strongly and half strongly suppress it. The separation between Metal and Wood signals is cleaner than in TN–H2, but still only 75% accurate. Since Air4H(2)T classifies Striker with 84% accuracy, it may be inferred that one or more of its other hidden nodes separate(s) the signals by some other criteria useful to the Striker nodes.

Negative cumulative sums (activations less than 0.5) were achieved by six of the twelve Air signal classes, in two different ways. The shortest signals (B1M, B1P, and B1W) simply decay so quickly that they fail to overcome the negative effective bias (see Figure 10.3.2.2.2-3(a)). This is identical to the way these signals were given positive sums by TN–H2. The other signals to

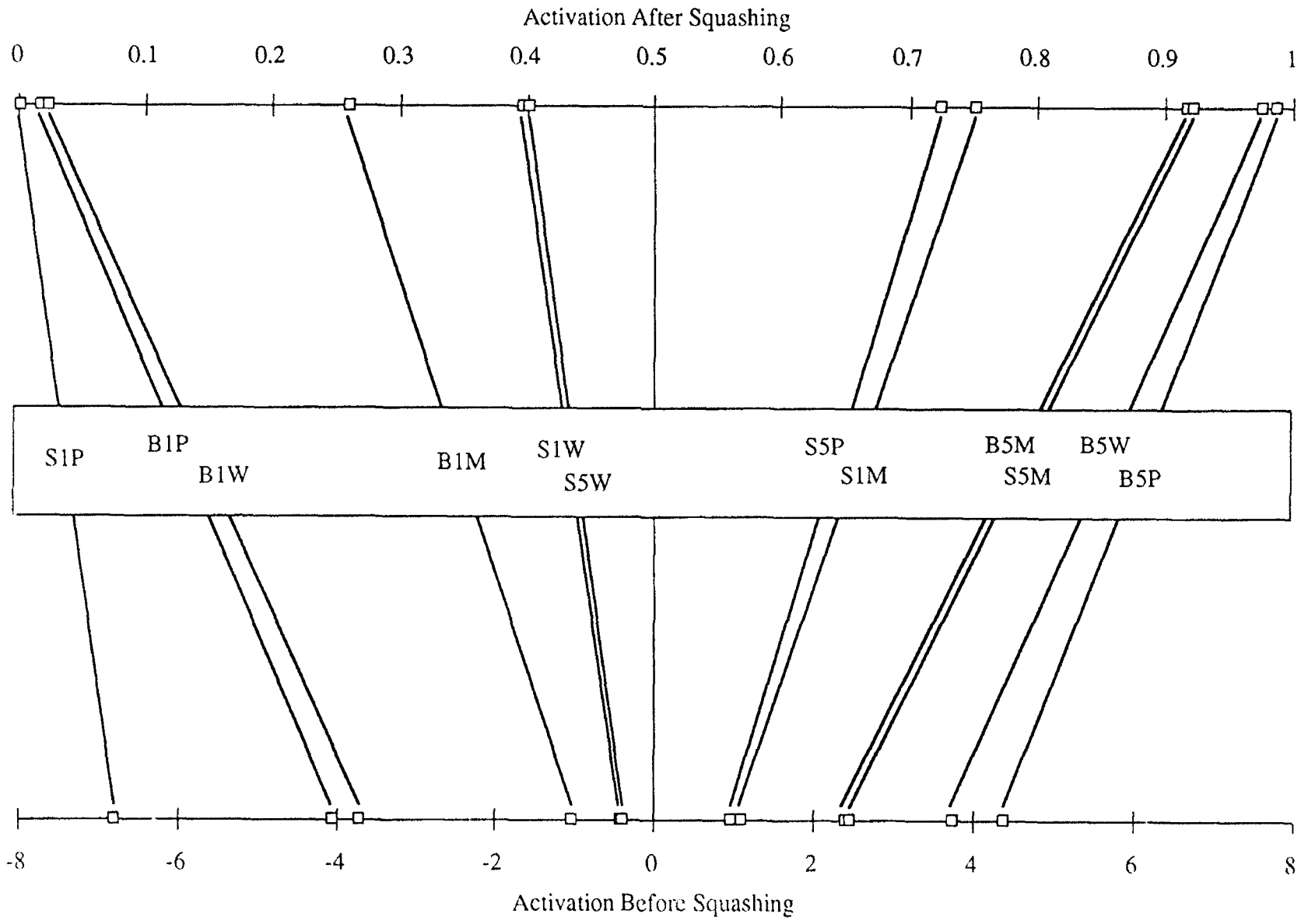


Figure 10.3.2.2.2-2 Hidden Node Air4H(2)T-H3 Activation for Instance Nine of Each Signal Class

Figure 3(a): B1P

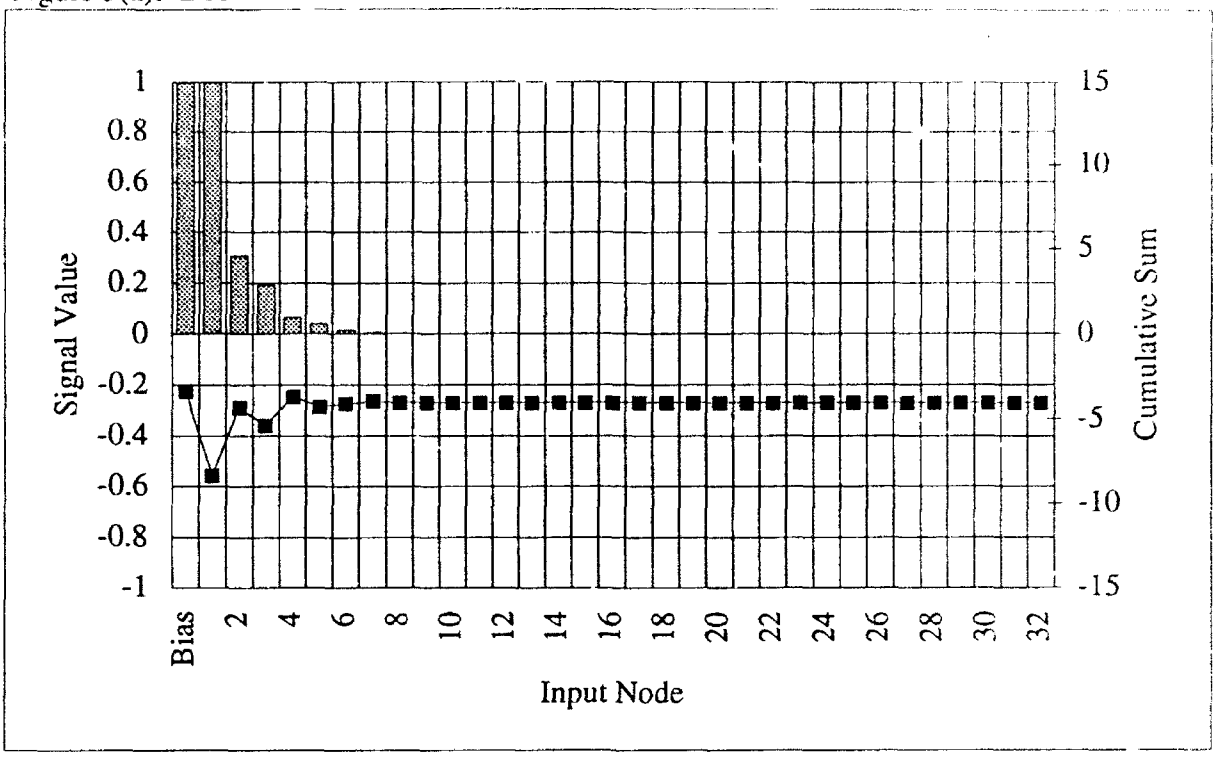


Figure 3(b): S1W

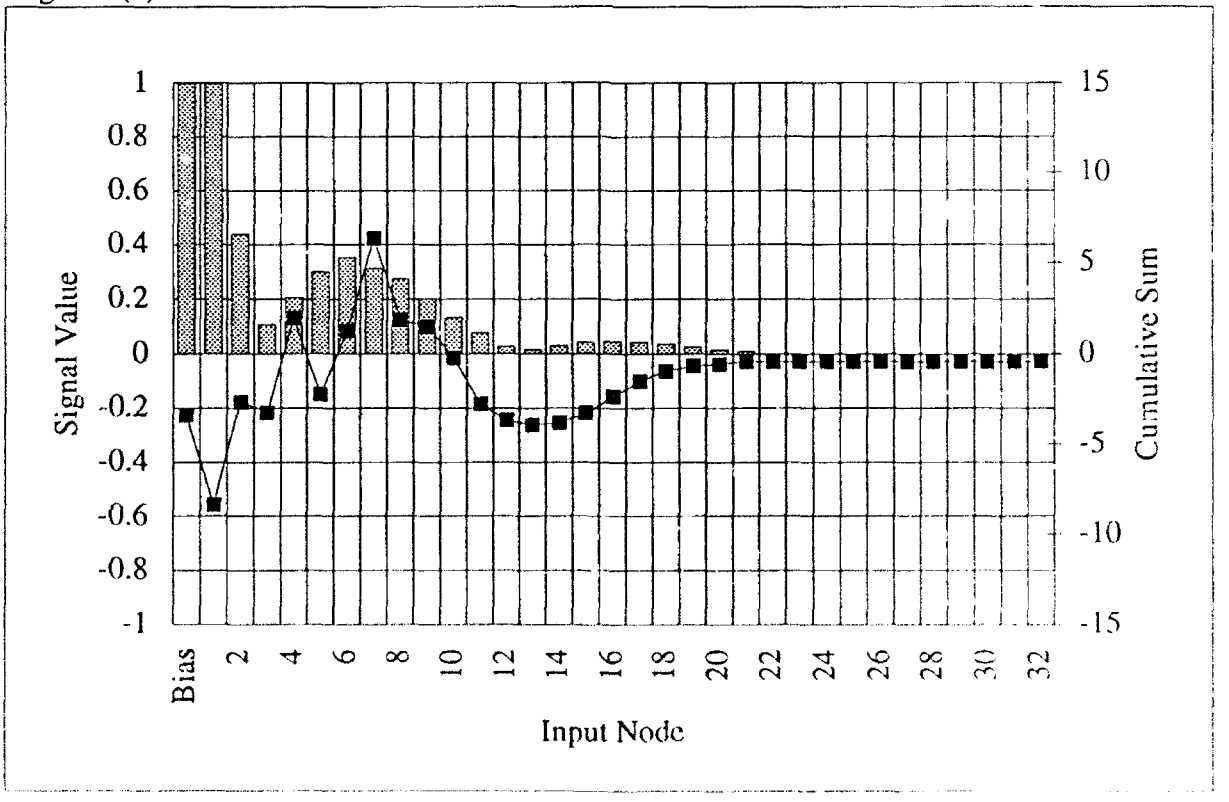


Figure 10.3.2.2.2-3 Cumulative Sum of Hidden Node Air4H(2)T-H3 for Instance Nine of Classes B1P and S1W

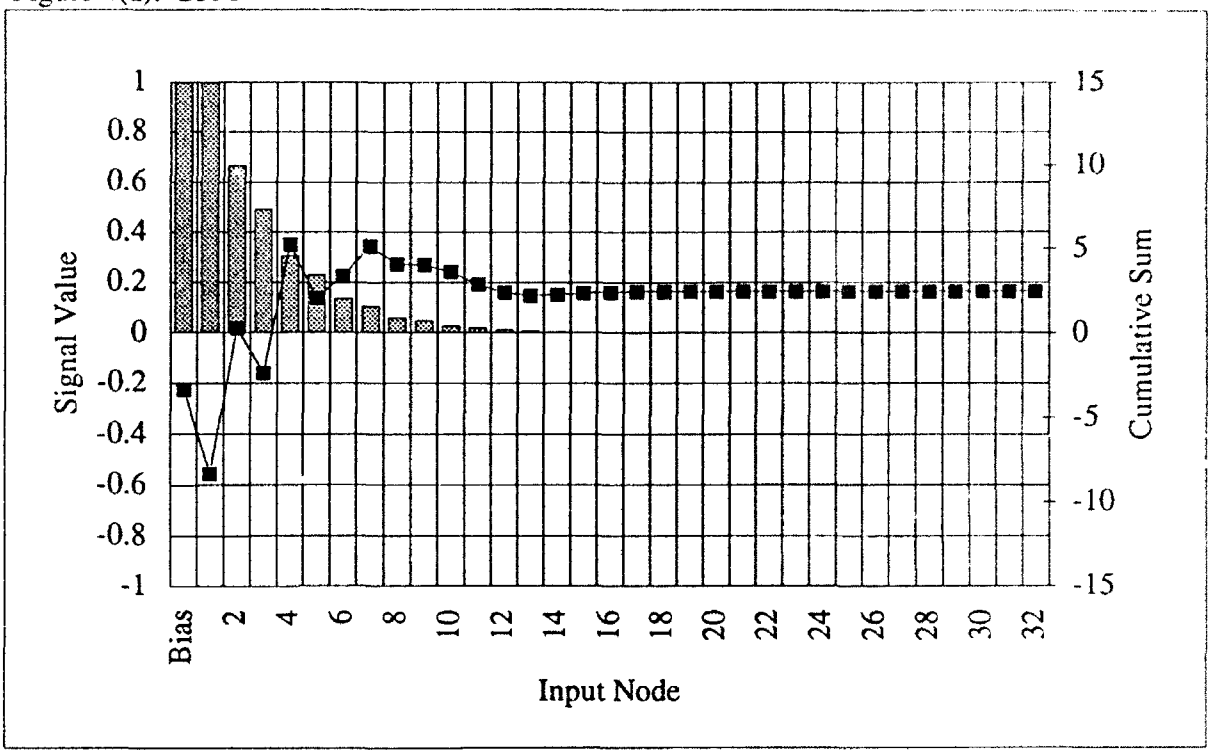
achieve negative cumulative sums were the boomerang signals (S1W, S1P, and S5W). These all had enough energy in T–I2 through T–I7 to overcome the negative effective bias, but were subsequently pulled back to negative cumulative sums by the second group of weights, and lacked the necessary energy in the third group to make the sum positive again (see Figure 10.3.2.2.2-3(b)). The remaining signals overcame the effective bias, and achieved net positive cumulative sums within the first weight group (T–I2 through T–I7), which were diminished by the negative second weight group (I8 through T–I13). Some signals (S5M and S5P) lacked the energy in this second region necessary to overcome the positive value achieved by the first weight group (see Figure 10.3.2.2.2-4(a)). The rest (B5M, B5P, B5W, and S1M) were actually pulled negative by the second weight set, then pulled back by the third to a final positive cumulative sum (see Figure 10.3.2.2.2-4(b)).

In summary, then, T–H3 uses information distributed throughout the signal to render its output for each signal. The weights fall naturally into three groups. The first group of weights is sensitive to the initial shape of the signal, providing the largest sum values for slowly decaying signals. The second group is negative, and reduces the sum for medium and long signals. The third group is positive, and counteracts somewhat the second group for long signals. The combined effect of all the groups is to produce high activations for long enveloped signals, and low activation for boomerang and short enveloped signals.

10.3.2.2.3 Comparison and Contrast of Hidden Nodes TN-H2 and T-H3

At the outset, the negative correlation between these two nodes suggested that they perform "opposite" functions. To some extent, this notion is reflected in the way the output layers use these two nodes. They are given weights of opposite sign and similar magnitude by the Metal, Wood, Brass and Steel nodes. However, while TN–H2 was ignored by Thickness nodes, T–H3 is used as a 5% detector. Both nodes use their input weights to extract information related to initial energy, decay, and duration of the signals. Both place emphasis on the first several inputs, gleaning from them a measure of how much signal energy is present, and how fast it is decaying. The hidden node TN–H2 essentially passes judgment on this information alone. Its weights are delicately balanced to yield a strong activation only for the three shortest signals, moderate activations for two of the three boomerang signals, and no or slight activation for the rest. Its function seems to be to detect only the quickest decaying signals, and the very slowest. Other signals are arbitrarily distributed between these extremes. The first several input weights of hidden

Figure 4(a): S5M



Signal Value — Cumulative Sum

Figure 4(b): B5M

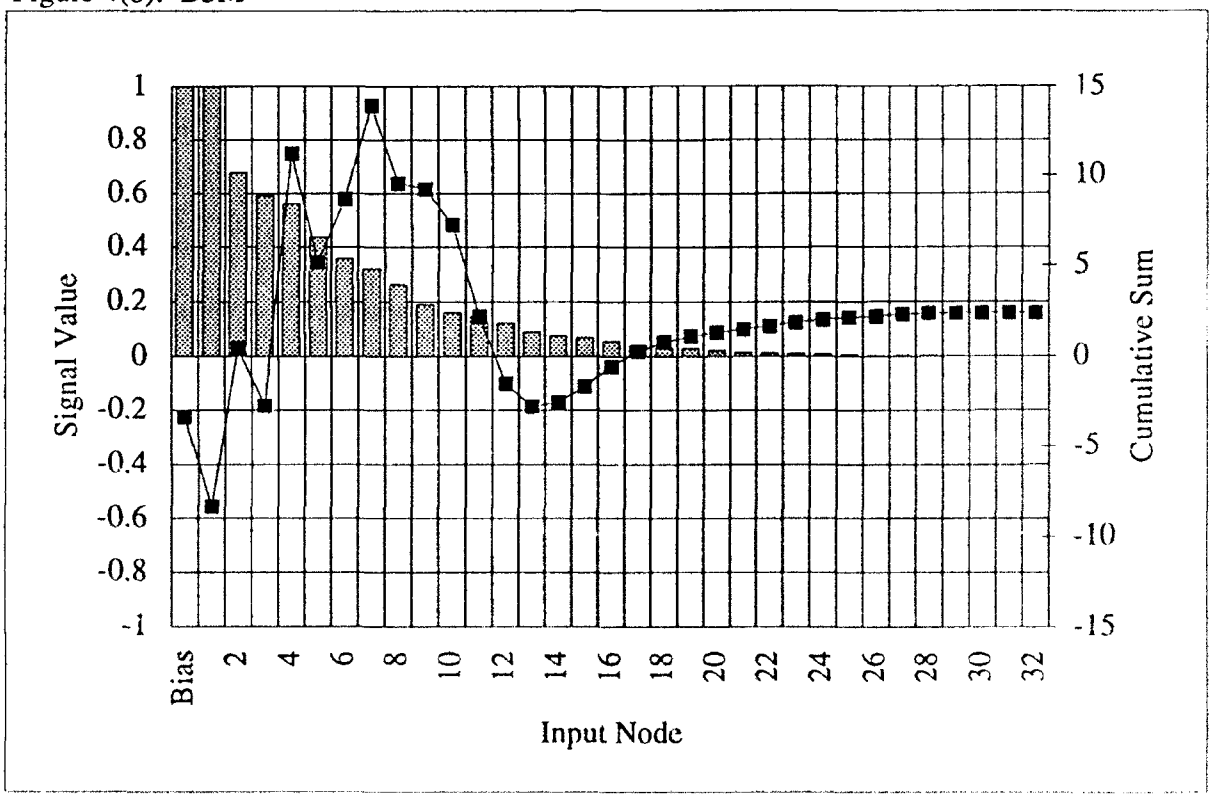


Figure 10.3.2.2.2-4 Cumulative Sum of Hidden Node Air4H(2)T-H3
for Instance Nine of Classes S5M and B5M

node T–H3 perform a very similar computation to TN–H2; the shortest signals are clearly identified. If energy is present in the second region, and not the third, the signal is identified as a boomerang signal. The last weights identify long ringing signals by their persistent low level tails.

The differences in the two algorithms developed by these nodes are most striking in the different ways they respond to S1P signals. Figure 10.3.2.2.3-1(a) shows the cumulative sum graph for an S1P signal applied to TN–H2. It is clear that TN–H2 is deactivated because of the first several large inputs. No second judgment is made by examining the total length of the signal, or its shape. By contrast, (see Figure 10.3.2.2.3-1(b)) the S1P signal initially activates T–H3 (the same judgment made by TN–H2), but this decision is reversed by the boomerang return energy. The S1P signal ultimately strongly suppresses T–H3, then, not because of its initial shape, but by its boomerang return and the lack of any later signal energy.

The functioning of T–H3 is more complex and more sophisticated than TN–H2, but at the same time less elegant and less general. The final output depends on a critical balance between almost all of the signal inputs. It is easy to see how the presence of added noise would disrupt this balance, particularly as the signal decays and the noise assumes greater relative value. It is particularly evident in the longest signals (see Figure 10.3.2.2.2-4(b)) that the cumulative sum wanders up and down a great deal before reaching its final value. This is partly a consequence of the large number of strong weights of either sign, and suggests that the algorithm developed by this node may render a value based on more "arbitrary" features of the particular signals included in the testing and training sets.

The simpler solution developed by TN–H2 uses much less of the signal information to determine its final activation. This restricts the node's ability to discriminate between signals, as there are cues in other portions of the signal which are ignored. Some of these cues are used by T–H3 to make a finer distinction between boomerang and long enveloped signals. On the other hand, TN–H2 classifies signals very similarly to T–H3, but performs this task much more simply and elegantly. It is more likely that TN–H2's classifies the signals using general features of the signal types, not arbitrary features of the signals in the training and testing sets.

The only difference in training between the two networks Air4H(2)T and Air4H(2)TN was that the latter was trained with noisy signals, while the former was not. The different weights that the hidden nodes TN–H2 and T–H3 developed are very suggestive about the effects of training noise.

Figure 1(a): S1P Applied to TN-H2

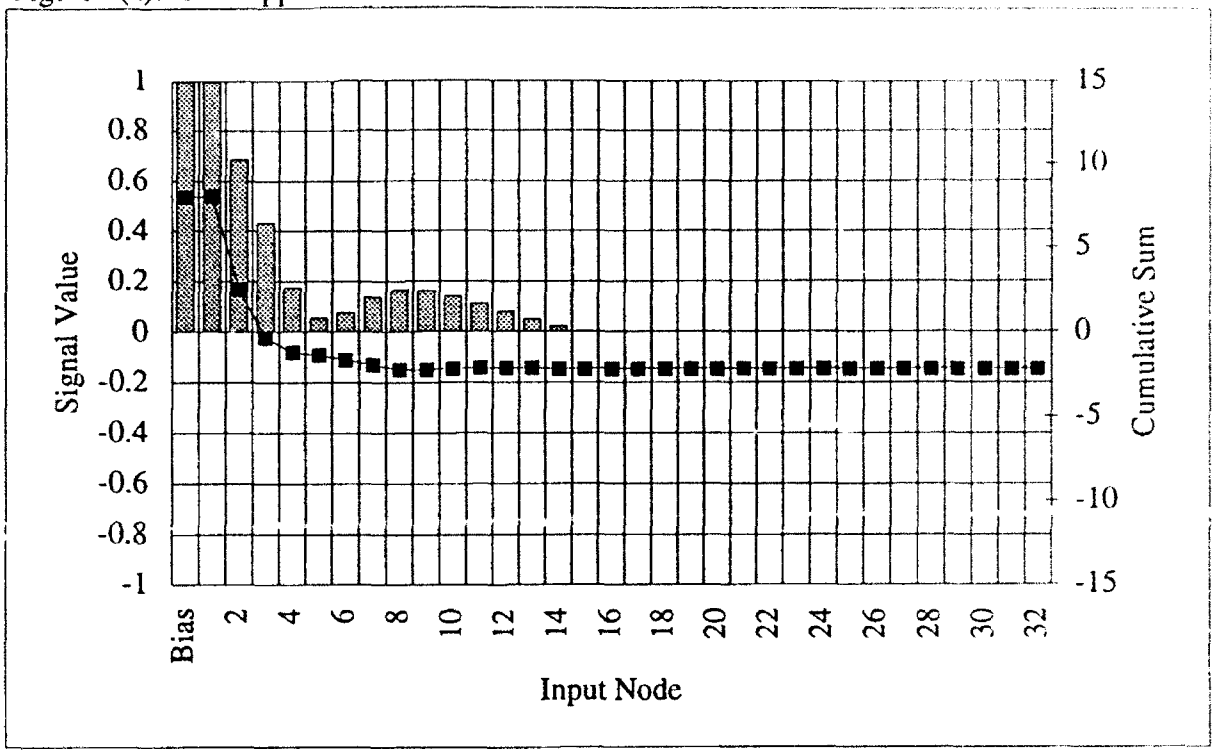


Figure 1(b): S1P Applied to T-H3

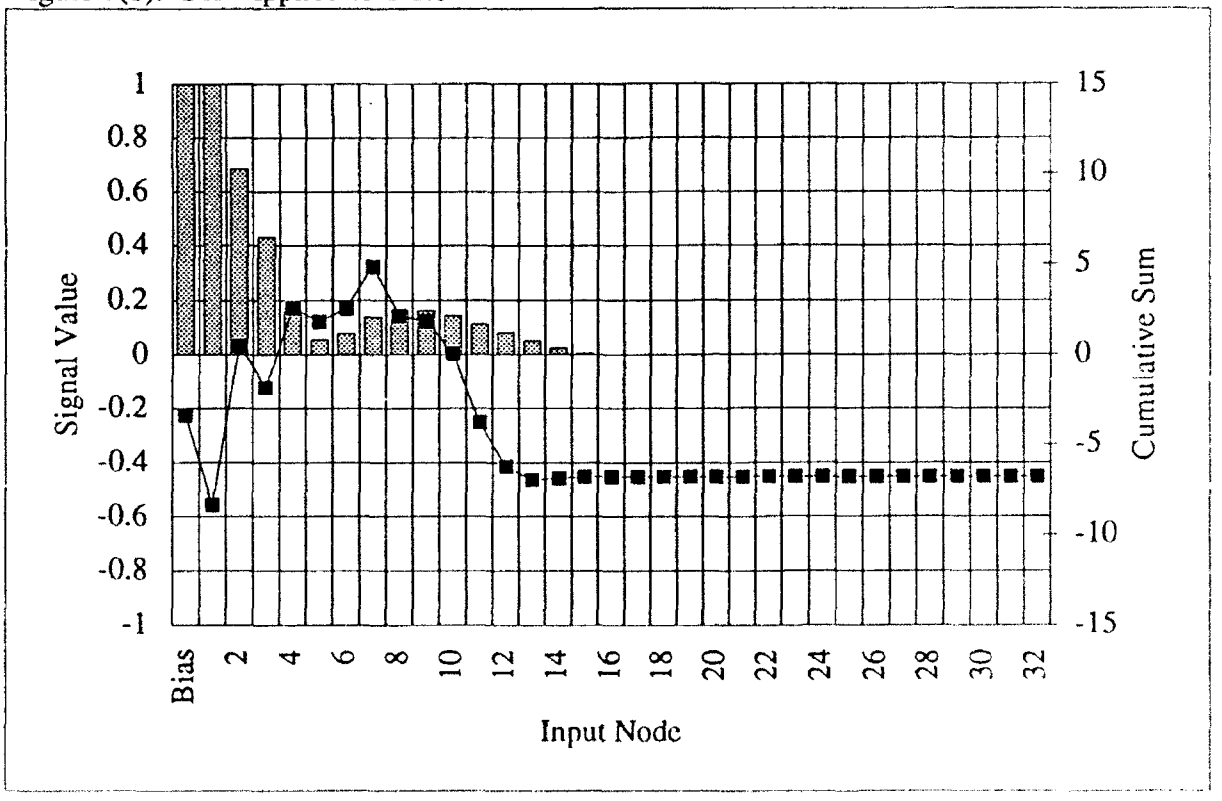


Figure 10.3.2.2.3-1 Cumulative Sum of Hidden Nodes Air4H(2)TN-H2 and Air4H(2)T-H3 for Instance Nine of Class S1P

The weights determined by TN–H2 use only the early portion of the signal, where the signal-to-noise ratio is largest. The addition of noise on training seems to have suppressed the use of weights in regions of the signal where the noise is more dominant. It is easy to see how this solution is more robust than that found by T–H3. The hidden node TN–H2 ignores those portions of the signal which are dominated by noise, and thus is able to process noisy inputs more consistently than its clean-trained cousin.

10.3.2.2.4 Comparison of TN-H2 and T-H3 to Best 1st and N4 1st Dimensions

In the case of the noise-trained hidden node (TN-H2), which was highly correlated with the Best first dimension and N4 first dimension (-0.81 and -0.83 respectively), we have seen a processing strategy extremely similar to that apparently used by the human subjects. Both the Best subjects and N4 alone placed the fastest-damping signals lowest on this dimension, as evidenced by the correlations with the two damping measures. It is reasonable to assume that the subjects were sensitive to these damping characteristics of the signals. TN-H2 made the same distinction using the same information. The weights of this hidden node reacted to fast-decaying signals with high activations, while producing low activations for long-decaying signals. Its weight structure was a simple, elegant means of measuring the decay characteristic of each signal.

The hidden node T-H3, trained without noise added to the signals, performed a calculation that may be considered an extension of that performed by TN-H2, although the calculation of T-H3 was considerably more complex. The strategy applied to the Brass 10% signals by T-H3 was the same as that of TN-H2 and the derived strategy of the subjects, i.e., the fastest decaying signal were separated from the others by their lack of energy beyond the first few inputs. Beyond these signals the strategies of T-H3 grew more complex and specific to particular signals. The long-decaying signals had to achieve their high activation using the third set of weights mentioned (T-I14 - T-I30), since they received large negative contributions from the second set of weights (T-I8 - T-I13). While it is not out of the question that subjects could have applied strategies as complex, the tools for deriving those strategies were not sensitive to such complexities. Keeping in mind that T-H3 was correlated with the first dimension of the Best solution at 0.94, one is led to believe that relatively complex processing was necessary to achieve such a close match to a dimension.

10.3.2.3 Analysis of Air4H(2)F-H2 and Air4H(2)FN-H2

Another pair of mutually correlated hidden nodes which also have substantial correlations with the Best and N4 first human dimensions are the frequency domain pair: Air4H(2)F-H2 (correlation -0.84) and Air4H(2)FN-H2 (correlation -0.82). In this case, the correlation between the two hidden nodes is +1.00, that is to say, perfect and positive.

10.3.2.3.1 F-H2 Analysis

The Thickness output nodes of Air4H(2)F place very large weight on Air4H(2)F-H2 (see Figure 10.3.2.3.1-1). In fact, it is safe to say that the only possible way a signal can overcome the substantial bias toward 5% thickness is by activating F-H2. By contrast, the other output nodes place relatively small weight on F-H2. Hence it may be concluded that F-H2's primary (and nearly exclusive) function is as a detector of 10% thickness. This task it performs perfectly, giving essentially 1.0 when a 10% signal is applied, and essentially 0.0 when a 5% signal is applied (see Figure 10.3.2.3.1-2).

Moving now to the weights connecting hidden node F-H2 to the input layer, several features stand out (see Figure 10.3.2.3.1-3). At the outset, one may notice that the bias term is very small. A substantial bias would imply that the node starts out strongly activated (or deactivated) and that its state is inverted by the presence of one type of signal (e.g. 5% or 10%). Because of the bias, the node would only have to recognize one type of signal to classify both types correctly. However, in the case of F-H2, the absence of a strong bias toward either signal type implies that the node actively detects each of the two types of signals it distinguishes. A glance at the weights reveals clearly that inputs I10 (4500 Hz) and I15 (7000 Hz) are the primary detectors of 5% signals (negative weights will tend to suppress the node), and that 10% signals are detected by a more distributed combination, with significant emphasis on inputs I6 (2500 Hz), I7 (3000 Hz), I19 (9000 Hz) and particularly I21 (10000 Hz).

To see how hidden node F-H2 detects 5% signals, consider Figure 10.3.2.3.1-4. The column plot in Figure 10.3.2.3.1-4(a) shows the frequency domain input of a B5W signal, while the superimposed line graph displays the cumulative sum of the hidden node F-H2 prior to the application of the transfer function. From the latter, it is clear that, although the final cumulative

Figure 1(a)

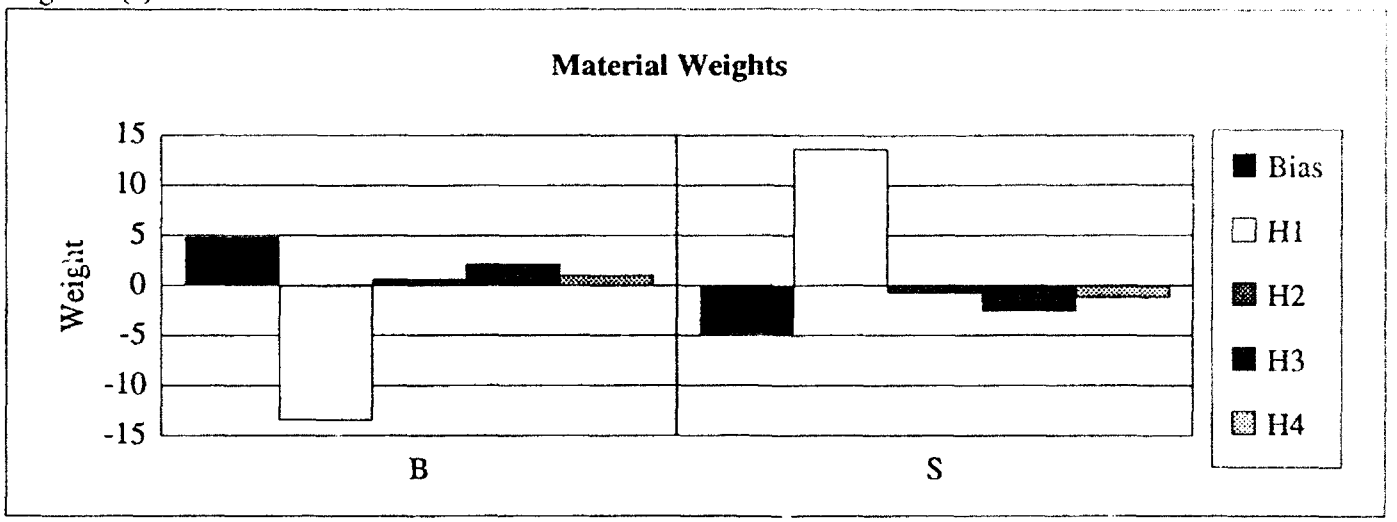


Figure 1(b)

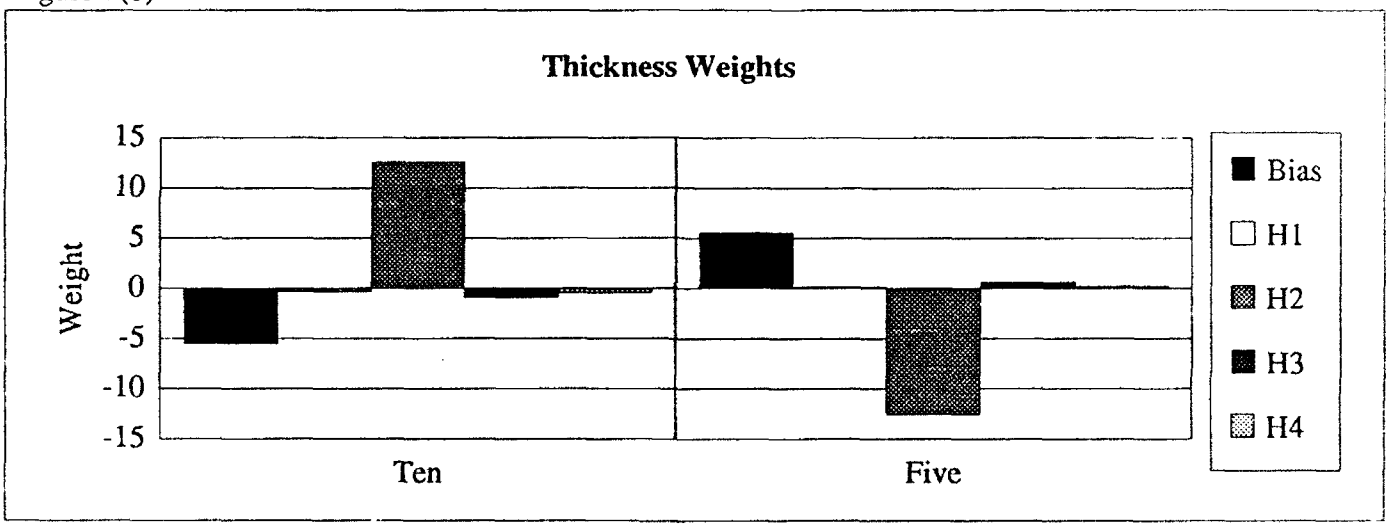


Figure 1(c)

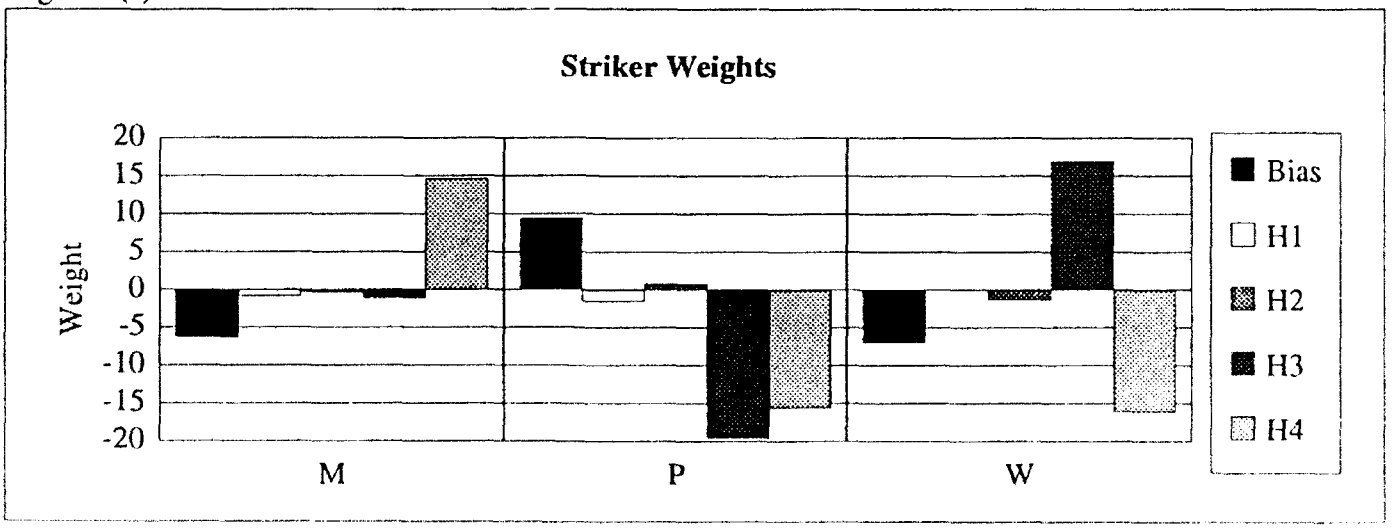


Figure 10.3.2.3.1-1 Weights on Hidden to Output Layer Connections in Air4H(2)F

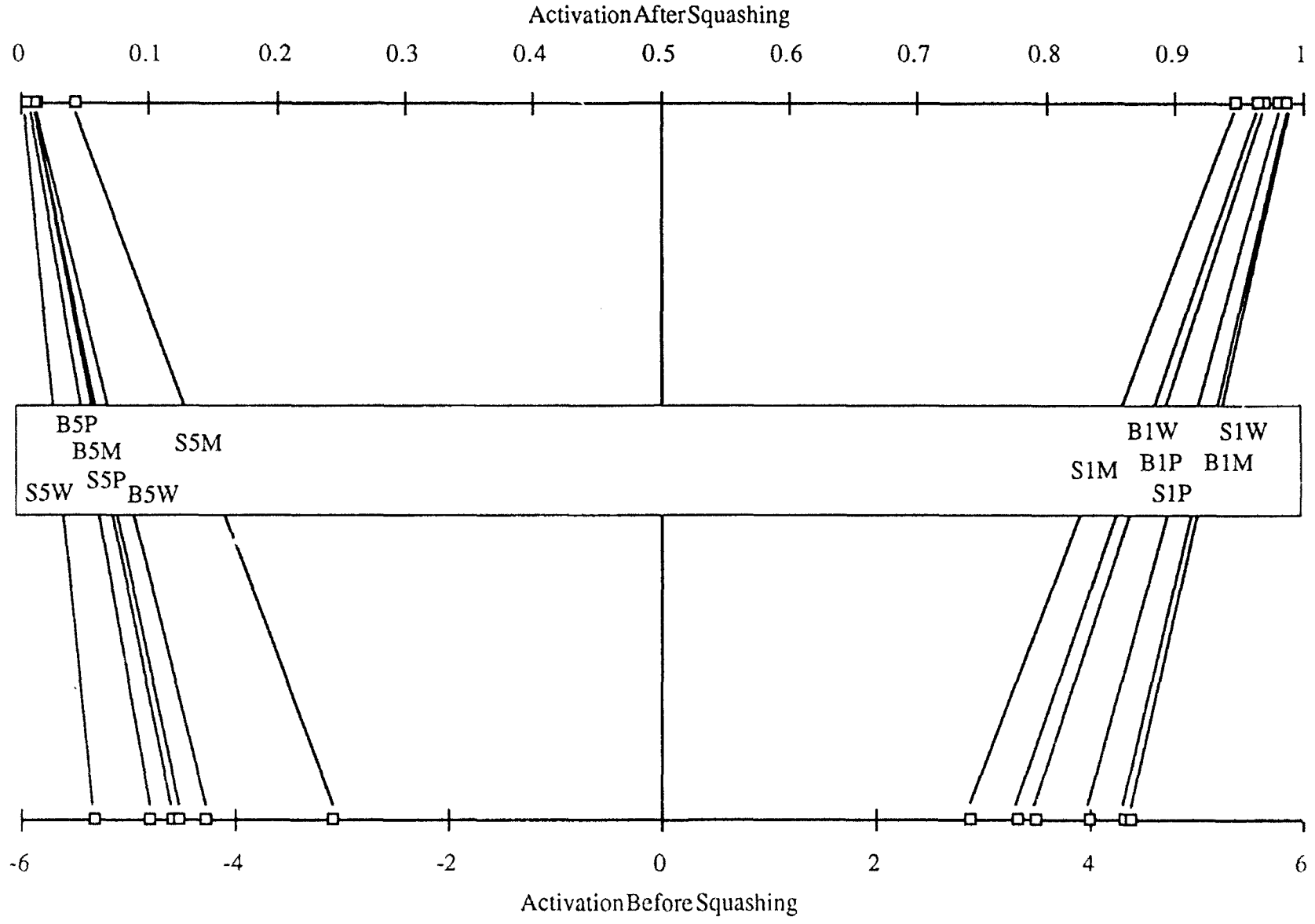


Figure 10.3.2.3.1-2 Hidden Node Air4H(2)F-H2 Activation for Instance Nine of Each Signal Class

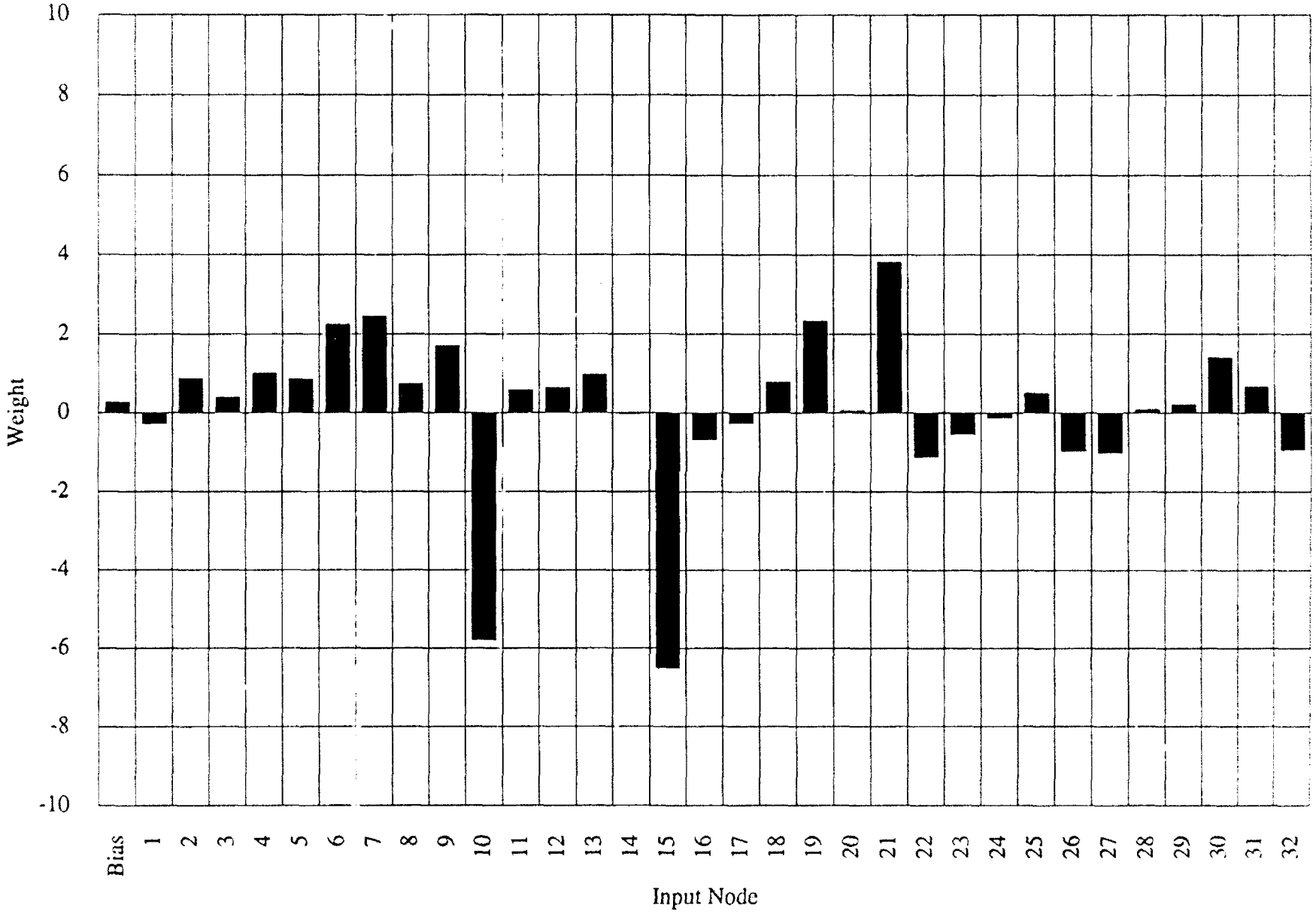
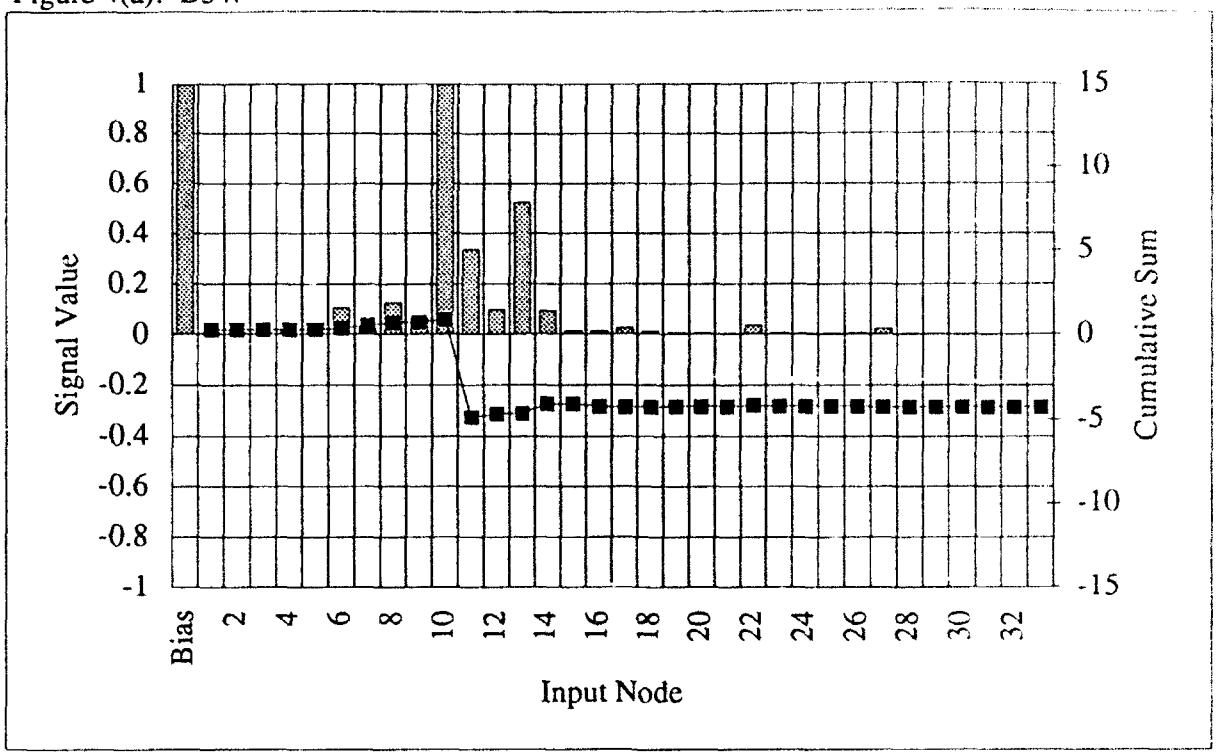


Figure 10.3.2.3.1-3 Weights on Input Layer to Hidden Node 2 Connections in Air4H(2)F

Figure 4(a): B5W



Signal Value — Cumulative Sum

Figure 4(b): S5M

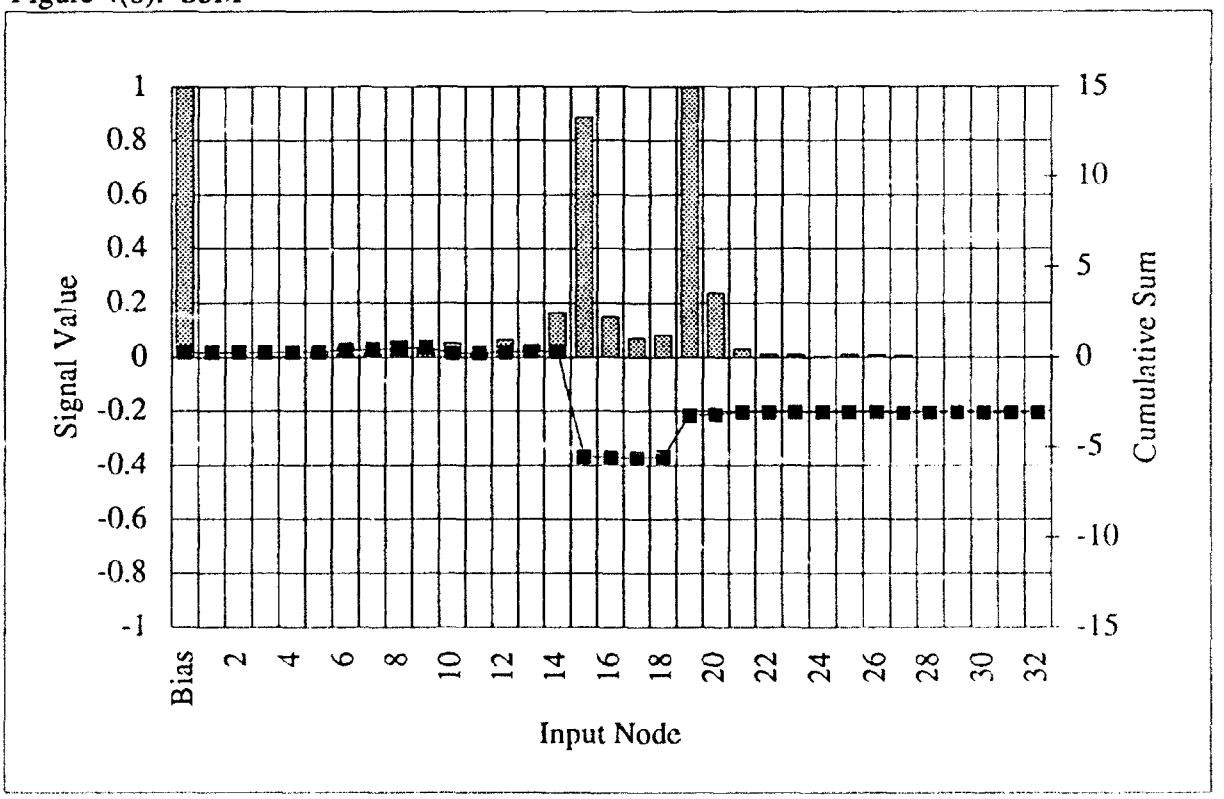


Figure 10.3.2.3.1-4 Cumulative Sum of Hidden Node Air4H(2)F-H2 for Instance Nine of Classes B5W and S5M

sum of the hidden node F-H2 is strongly negative (suppressed), it would be positive without the large amount of signal energy present in input I10 (4500 Hz). The precipitous jump in the output occurring at this input is both necessary and sufficient to classify this signal as 5%. Similar graphs plotted for the other Brass 5% signals show that they suppress hidden node F-H2 even more strongly on the strength of this input.

Figure 10.3.2.3.1-4(b) shows a corresponding plot for Steel 5% signals, in particular, an instance of S5M. Clearly, a similar situation exists here; the hidden node is suppressed in this case by the large signal input I15 (7000 Hz). Were it not for this input (and the corresponding negative weight), the cumulative sum would be forced positive by the signal energy present in input I19 (9000 Hz). The cumulative sum resulting from the application of the other Steel 5% signals follows the same pattern, and the output from F-H2 is even more strongly suppressed by them.

The classification of 10% signals performed by F-H2 is slightly more complex. Shown in Figure 10.3.2.3.1-5(a) is an instance of B1W and the corresponding cumulative sum obtained by F-H2. The positive final value of the cumulative sum results from a combination of large amounts of signal energy in inputs I13 (6000 Hz) and I21 (10000 Hz), and more modest energy in inputs I1 through I12 (0 - 5500 Hz). The large negative weights on inputs I10 (4500 Hz) and I15 (7000 Hz) reduce the cumulative sum, but the signal energy in these inputs is insufficient to suppress hidden node F-H2. This is again typical of the other Brass 10% signals.

The Steel 10% signals show the largest variation in the shapes of their inputs (see Figures 10.3.2.3.1-5(b) and 10.3.2.3.1-5(c)). Nevertheless, they have one common feature: the maximum signal energy is found in input I19 (9000 Hz). The high positive weight on this input is enough to activate hidden node F-H2. In the case of S1M, this is the only significant contribution to the cumulative sum, as shown in Figure 10.3.2.3.1-5(b). The input patterns of the Steel 10%, Plastic and Wood signals are similar to each other, and more complex. The input and cumulative sum for an S1P signal is shown in Figure 10.3.2.3.1-5(c). In this signal, there is significant energy in inputs I10 (4500 Hz) and I15 (7000 Hz). While the energy present in I19 is still necessary for strong activation, it is not sufficient, due to the negative contributions in these two inputs. The large negative jumps caused by these two "5%-like" inputs, especially the input I10, are counteracted by the wide distribution of signal energy in inputs I1 through I9 (0 - 4000 Hz), and I11 through I13 (5000 - 6000 kHz).

Figure 5(a): B1W

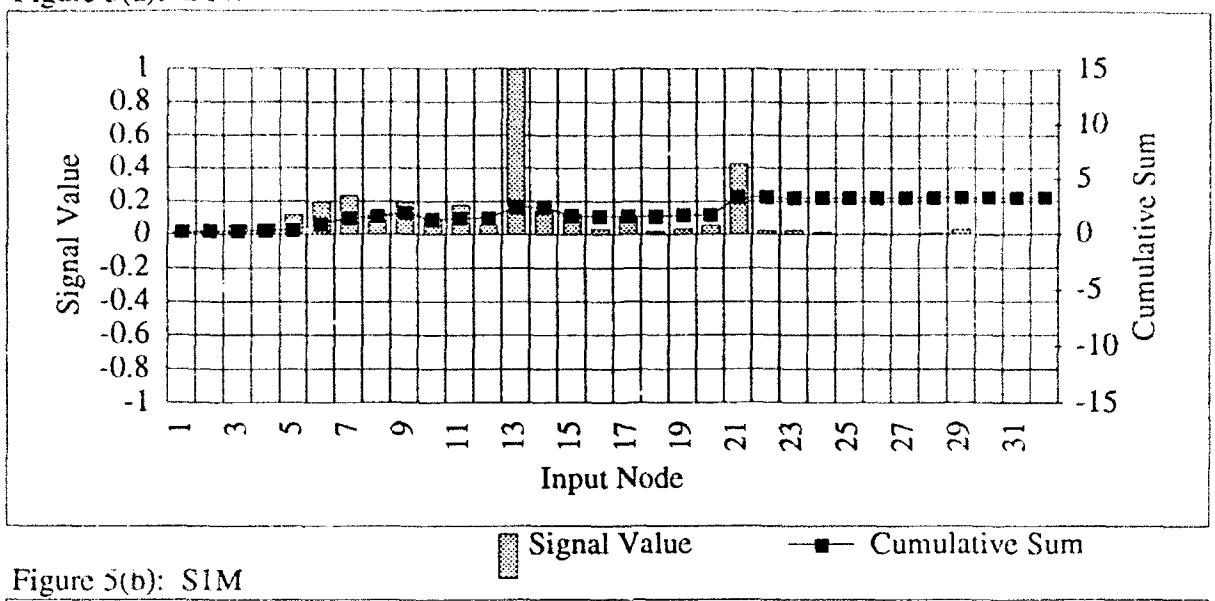


Figure 5(b): S1M

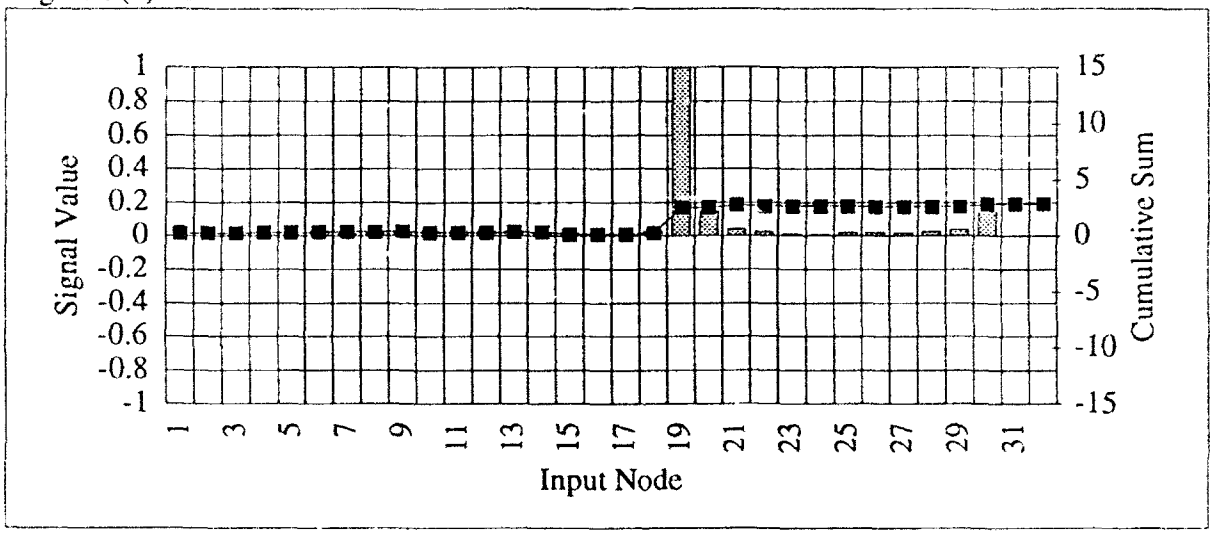


Figure 5(c): S1P

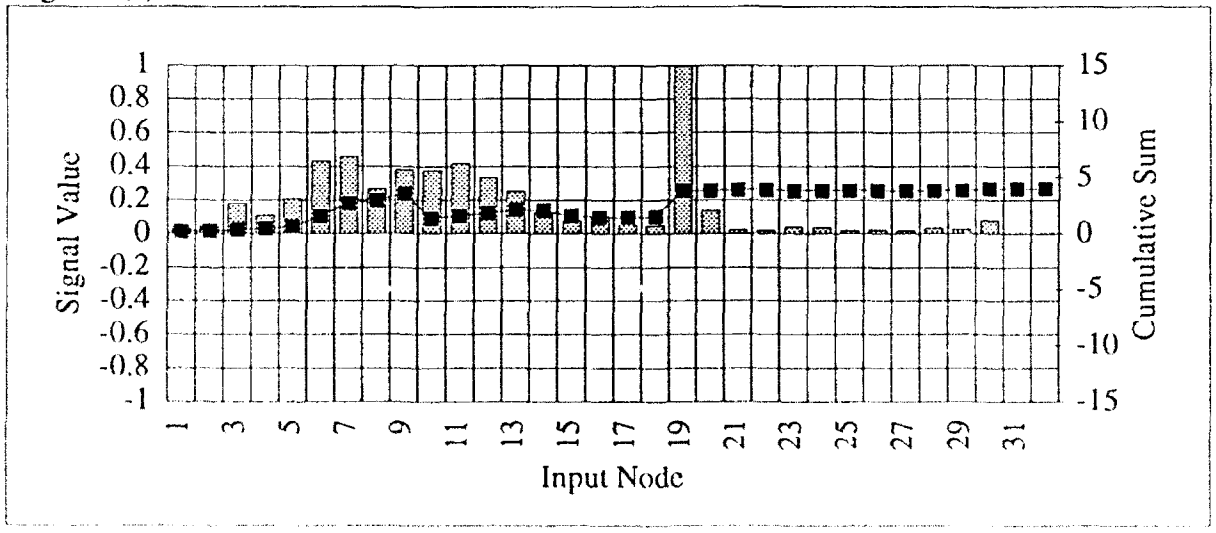


Figure 10.3.2.3.1-5 Cumulative Sum of Hidden Node Air4H(2)F-H2 for Instance Nine of Classes B1W, S1M, and S1P

To summarize, hidden node F-H2 classifies signals according to the Thickness of the target. The inputs which are most important are inputs I10 (4500 Hz) and I15 (7000 Hz), which are very pronounced in the Brass 5% and Steel 5% signals, respectively. While there is substantial energy in these inputs in some of the 10% signals, it does not suppress the node, due to a broad distribution of moderate energy in neighboring (positively weighted) inputs, and large amounts of energy in inputs I19 (9000 Hz) and I21 (10000 Hz) in the Steel 10% and Brass 10% signals, respectively.

10.3.2.3.2 FN-H2 Analysis

The other hidden node in this pair belongs to the related network Air4H(2)FN, similar in all respects to the network discussed above, save that it was trained with signals to which noise had been added. A comparison of Figures 10.3.2.3.1-1 and 10.3.2.3.2-1, which show the output weights for networks Air4H(2)F and Air4H(2)FN, respectively, reveals that these two networks weight hidden node H2 almost identically. From the output layer, it is therefore clear that FN-H2 is also a detector of 10% signals. This is not too surprising since the correlation between the pair F-H2 and FN-H2 was +1.00. It was known at the outset that these nodes sorted the signals into the same order. It is possible, however, for the two nodes to develop very different means of performing this classification. In this case, however, differences between the input weight structures of the two nodes are completely inconsequential (see Figures 10.3.2.3.1-3 and 10.3.2.3.2-2). Many of the smaller weights differ noticeably between the two hidden nodes, but the large, influential weights are virtually identical in both. It is interesting that training with noise had a large effect on the weights developed in the time domain Air 4 hidden node networks, but very little effect on the hidden node Air4H(2)F-H2. This may have some bearing on the fact that among the Air 4 hidden node networks in frequency domain, the clean-trained network actually performed better than the network trained with noisy signals.

10.3.2.3.3 Comparison of F-H2 and FN-H2 to Best 1st and N4 1st Dimensions

As discussed earlier, the two first dimensions have both time domain and frequency domain explanations which were demonstrated by correlations with signal measures in both domains, and by listening. The two frequency domain hidden nodes reflect some of the same processing strategies that were found earlier, namely in the relationships between these dimensions and the standard deviation and curve fit frequency. Both the standard deviations of the signals and the

Figure 1(a)

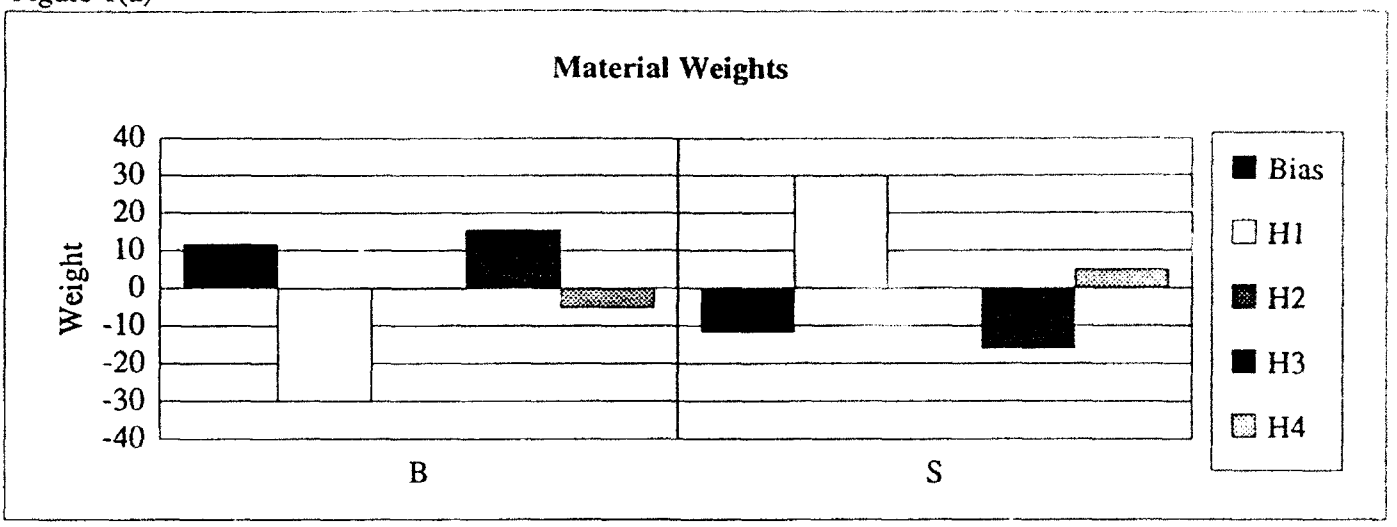


Figure 1(b)

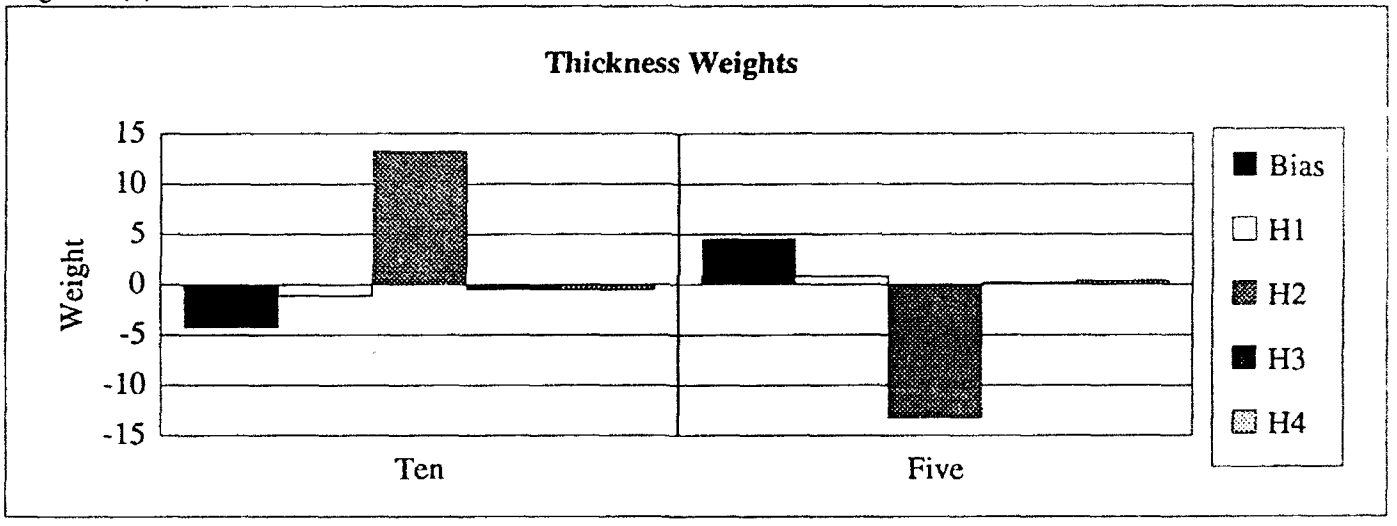


Figure 1(c)

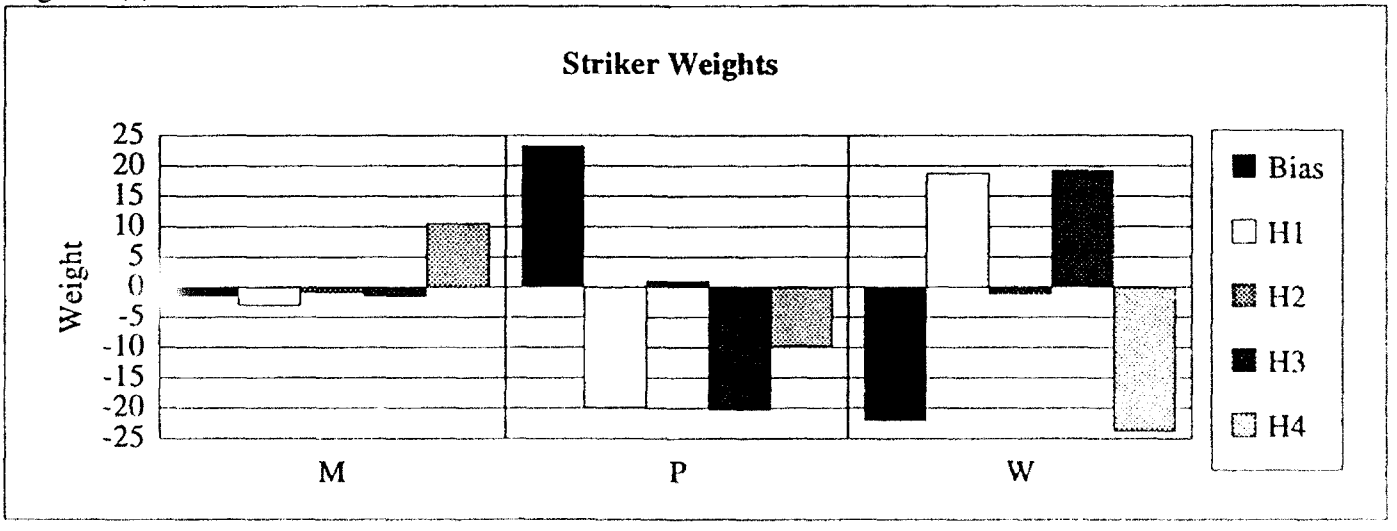


Figure 10.3.2.3.2-1 Weights on Hidden to Output Layer Connections in Air4H(2)FN

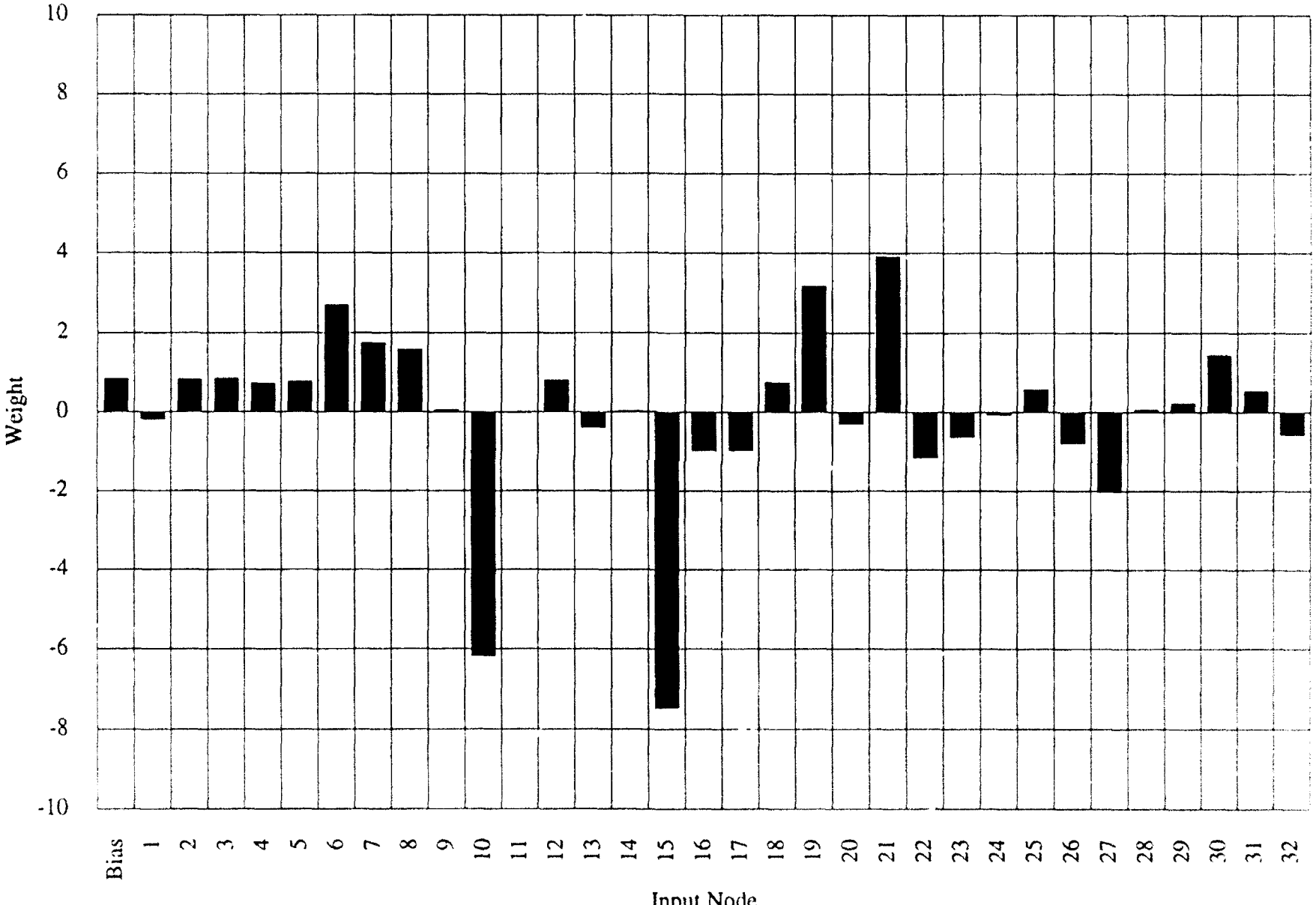


Figure 10.3.2.3.2-2 Weights on Input Layer to Hidden Node 2 Connections in Air4H(2)FN

frequencies of their most persistent sine waves (found in the curve fit solutions) are negatively correlated with both dimensions. The weights on both hidden nodes reflect the relationships with standard deviation and frequency, indicating that the nodes are sensitive to signal features similar to those to which the subjects appeared to be sensitive (in some combination with the time domain feature of decay) on these dimensions.

Viewing the weights of F-H2 gives strong indications of the sensitivity of this node to both the standard deviation and frequency of the ringing portion of the signals. The node gives strong activation for 10% signals, and does so using the large positive weights in bins 6 and 7, and those in 19 and 20. The node is suppressed by signal energy in bins 10 and 15, which are closer together than the bins needed to excite the node. Assuming signals provide energy in both areas in order to excite or suppress the node, signals that suppress the node have a smaller distribution of energy than signals which excite the node. This is in keeping with the negative correlation between standard deviation and the two first dimensions for Air signals. Examination of the frequency distributions of the signals reveals that those signals of 10% thickness with Plastic and Wood strikers have strong low frequency components spread over several bins as well as peak frequencies at bins 19 or 21. Signals of 5% thickness do not have substantial frequency components at these extremes.

The node also tends to activate strongly for signals with high frequencies of their most persistent sine wave component. This signal measure is only concerned with the frequency component which persists the longest in the signal and is the portion of the signal which a listener may describe as the ringing portion. We assume that this frequency component dominates the spectrum. We then note that Steel 10% signals ring at bin 19 (which starts at 9.0 kHz), while Steel 5% signals ring at bin 15 (7.0 kHz). The Brass 10% Metal signal rings at bin 21 (10 kHz), the other Brass 10% signals at bin 13 (6.0 kHz), while Brass 5% signals ring at bin 10 (4.5 kHz). When only the ringing frequency is considered, the node activates strongly for signals of high frequency, which tend to be the 10% signals.

10.3.2.4 Summary of Hidden Node Processing

In summary, the two frequency domain hidden nodes were trained to give high activations for signals of large standard deviation and high ringing frequency. These two signal characteristics were also the primary frequency domain features by which the signals on the first human

dimensions of each scaling solution are sorted. They were audible to the listener as described above, and presumably were part of the subjects' processing strategies. The node of course must identify the frequencies exactly, while the subjects were free to apply a less restrictive rule. The time domain hidden nodes described previously also found a signal feature by which the first dimensions are sorted, a feature related to the damping characteristics of the signals. Both time and frequency domain nodes appeared to be applying signal processing strategies which are closely related to those of the subjects on the first dimensions of each scaling solution.

10.3.3 Best Second and N4 Third Dimensions

These dimensions are quite highly correlated at -0.93, yet there are important differences in the breakdown of signals. The high correlation is due to strong similarities in the extremities of the dimensions. In particular the Brass 5% signals are grouped at opposite extremes of each dimension. At the other extremes of each dimension are S1M and S5M. The remaining signals are distributed between the extremes in a somewhat different manner for each dimension. The Best second dimension places S5P with the extreme Steel signals, but has all other signals in a relatively tight group in the middle of the dimension, with no apparent ordering by parameter.

The N4 third dimension is arranged differently in the middle. This dimension divided the signals by Material with no overlap. All Steel signals are lower on the dimension than any Brass signals, although the nearest two signals of different Material are very close. The signal feature represented by this dimension was presumably used by N4 to make Material judgments, which this subject did with approximately the same high performance as the other two subjects in the Best solution (0.86). No other dimension of N4 differentiates Material. The third dimension of the Best solution has the signals separated by Material with one exception, yet the values of the signals are different enough from the N4 third dimension to prevent a significant correlation.

10.3.3.1 Dimensions Analysis

The two dimensions in question are highly correlated only with two frequency domain statistics, the mean and mode. These measure the location of the "center" of the frequency distribution, one by taking an arithmetic mean and the other by identifying the single strongest frequency. In practice on these signals the two are very similar. The correlations indicate that signals high on the

Best second dimension have high mean and modal frequencies, while signals high on the N4 third dimension have low mean and modal frequencies.

Listening confirms these relationships. The placement of the Brass 5% signals low on the Best second dimension is accounted for aurally by the low frequency of the ringing portion of these signals. This low frequency component distinguishes the Brass 5% class from all other signal classes which ring for the same duration as the Brass 5% signals. This relationship accounts for the strong correlations between the dimension and the mean and modal frequency statistics. This frequency characteristic does not extend, however, to signals other than Brass 5%. That is, the remaining signals taken by themselves do not show correlation with the mean frequency, nor with other signal statistics. Listening reveals some characteristics of this group of remaining signals. Three signals are very high on the dimension: S1M, S5M, and S5P. These signals share the characteristics, relative to the six remaining signals, of having long ringing portions and very little impact sound distinct from the beginning of the ring. The remaining six signals, which are relatively close together near the middle of this dimension, have distinct impacts followed by either a vibrato ringing portion in the remaining three Steel signals or virtually no ringing portion at all in the Brass 10% signals.

Listening to the N4 third dimension leads to similar observations. The mean has a high negative correlation with this dimension primarily due to the placement of the Brass 5% signals high on the dimension. These signals have considerably lower mean frequencies than all other signals, and this effect is easy to hear when listening to the signals ordered on this dimension. The mean would not appear to be highly correlated with the dimension if the Brass 5% signals were not considered. Listening suggests that the subject was using a combination of mean frequency and ringing characteristics on this dimension. Note that the material of the target is perfectly separated on this dimension (although the difference between B1M and S1W is very small). The three Brass 10% signals are very highly damped. Subjectively, this serves to diminish the perception of high frequency content in these signals. While for their relatively brief duration they actually have a fairly high mean frequency, their damping tends to mask this content. This suggests that the subjects placed these signals lower than the Steel signals due to a perceived lack of high frequencies.

Mean is a reasonably good predictor of the signals' values on the Best second dimension:

R^2(adj) = 54.3%
Mean p = 0.0037

Mode is almost as good, but both are good predictors only in their ability to discriminate Brass 5% signals from all others.

As expected, mean is also the best predictor for N4:

R^2(adj) = 54.6%
Mean p = 0.0036

Although in this case, low frequency slope made a significant addition to the regression:

R^2(adj) = 68.6%
Mean p = 0.0015
Low Frequency Slope p = 0.0445

Low frequency slope is used to discriminate the Brass 10% signals from the remaining signals, as shown in Figure 10.3.3.1-1. Brass 10% signals have a higher slope, indicating that they have a sharper cutoff of low frequencies, presumably related to their rapid damping characteristic.

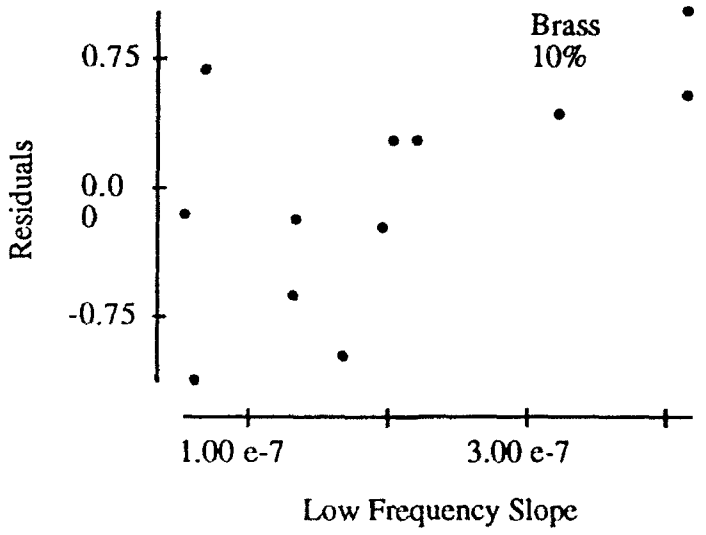


Figure 10.3.3.1-1 Low Frequency Slope vs. Regression Residuals

10.3.3.2 Analysis of Air4H(2)T-H2 and Air4H(2)TN-H3

This pair of nodes is of particular interest due to correlations with the human subject dimensions Best-2 and N4-3. There is also a large, negative correlation between the two nodes themselves. The parent networks of these nodes, Air4H(2)TN and Air4H(2)T, were trained from identical initial conditions, with and without training noise, respectively. It is desirable to begin by studying the node from the network trained with noise, since it presents a simpler input weight structure. For brevity, within the following discussion, these two networks will again be referred to simply as TN and T.

10.3.3.2.1 TN-H3 Analysis

Among the output nodes, TN-H3 is used very heavily as a detector both of Brass and 5% signals (see Figure 10.3.1-2). It does not serve to detect either Metal or Wood signals, and although the Plastic output node (TN-P) places positive weight upon it, it is doubtful that TN-P performs a useful computation. Thus, the node is used to determine Material and Thickness, but not Striker.

The input weights for TN-H3 are shown in Figure 10.3.3.2.1-1. The large negative bias can be overcome by sufficient energy in the range TN-I2 through TN-I7. Outside this range the product of the decaying signal inputs with the small weights is too small to influence the final state of the node significantly. Rapidly decaying signals cannot overcome the bias, and thus remain strongly negative, while the activation resulting from longer signals is less negative or even positive, depending on the exact distribution of energy. The general shape and behavior of the input weights is very similar to that of another node in the same network, TN-H2. It turns out that there are some interesting similarities between these two nodes, which will be discussed later.

The responses of TN-H3 to instance nine of the various signals are shown in Figure 10.3.3.2.1-2. Before applying the transfer function to the outputs (lower axis plot), the largest division between any two signals is the gap of approximately 5.3, between classes S1M and B5M. This break is the only one to which any significance can with confidence be ascribed; it divides the Brass 5% signals from the rest. The upper axis plot shows the final result after application of the transfer function. Since only Brass 5% signals activate the node, it is a perfectly accurate detector of these signals. This is consistent with the heavy weights placed upon TN-H3 by the output nodes TN-B and TN-Five. This fact, in consideration of the weights placed upon the other hidden nodes, also

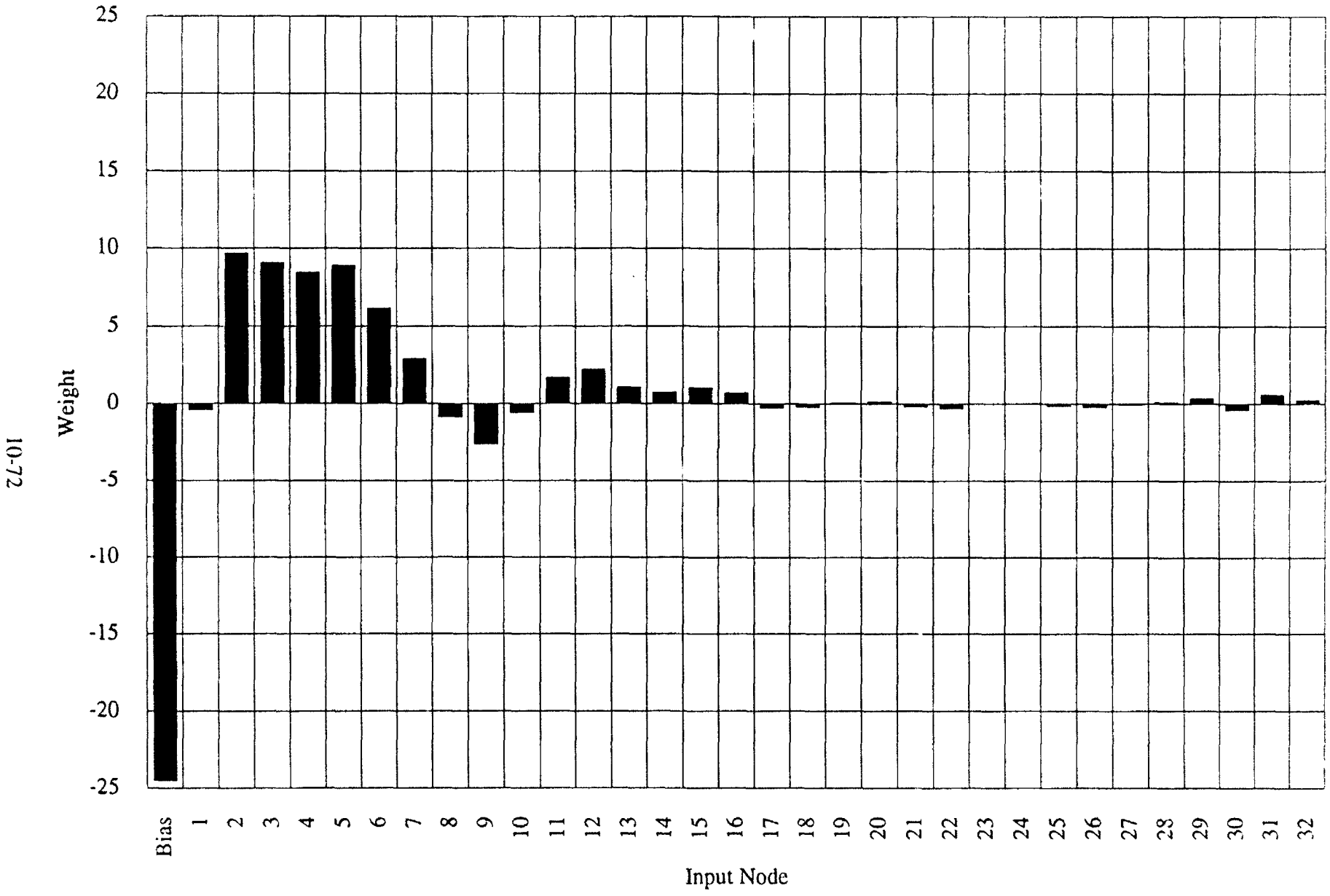


Figure 10.3.3.2.1-1 Weights on Input Layer to Hidden Node 3 Connections in Air4H(2)TN

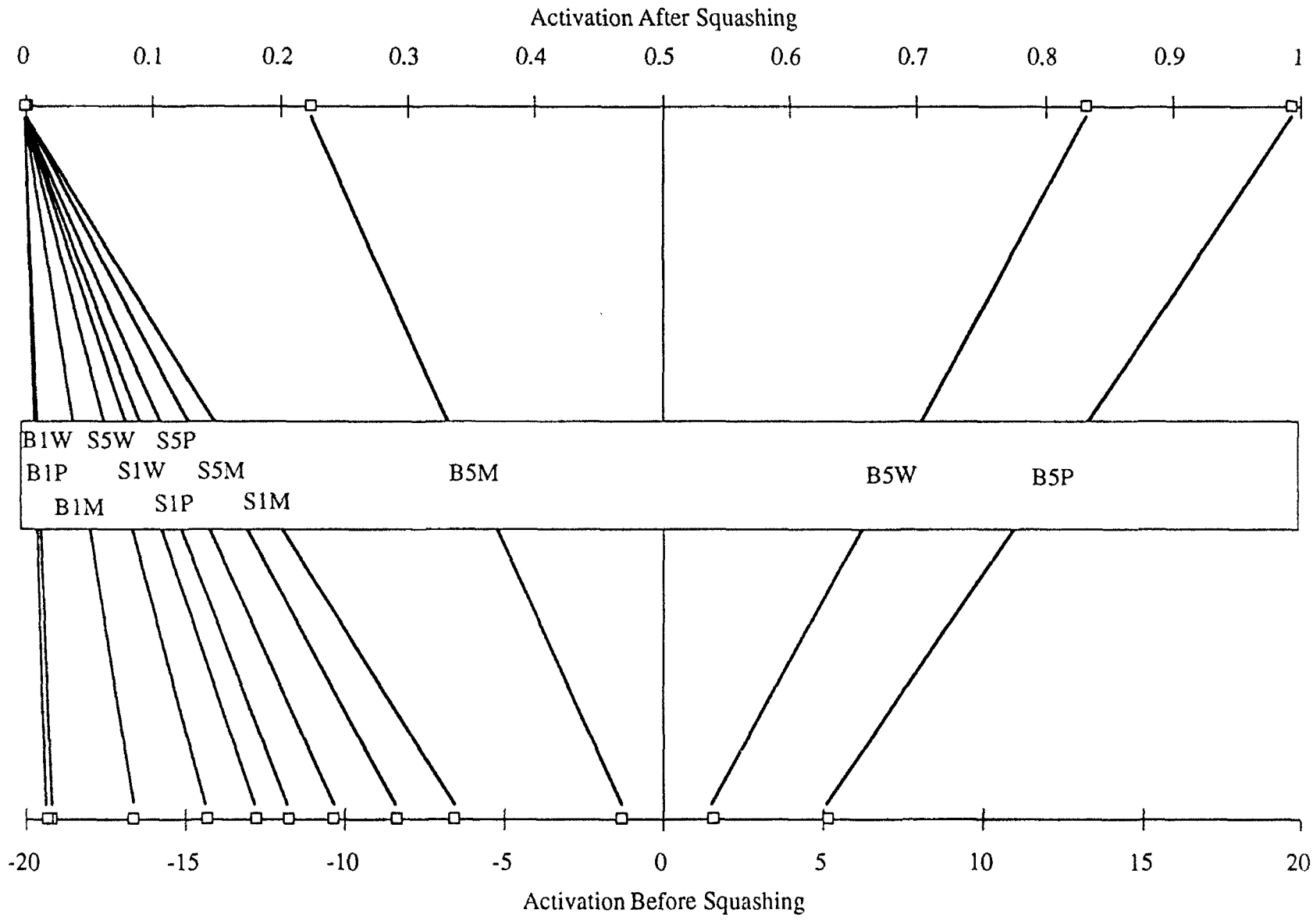


Figure 10.3.3.2.1-2 Hidden Node Air4H(2)TN-H3 Activation for Instance Nine of Each Signal Class

implicates hidden node TN-H1 as a detector of Brass 10% signals. Cumulative sum graphs for all signals were examined. From this it is obvious that the cumulative sums for TN-H3 yielded virtually identical (inversion notwithstanding) patterns to those for TN-H2.

10.3.3.2.2 T-H2 Analysis

The cleanly trained counterpart of TN-H3 is the hidden node T-H2. The two nodes have a correlation of -0.99, suggestive that their algorithms sort the signals into opposite orders. The Material and Thickness weights of the network are consistent with the negative correlation; the node is used moderately as a detector of Steel 10% signals. In contrast to TN-H3, however, T-H2 is used as a detector of Metal, and a rejector of Wood and Plastic Strikers.

As might be expected in a network trained with clean signals, the input weight structure of T-H2 is much more complex than that of TN-H3 described above. The input weights of T-H2 are shown in Figure 10.3.3.2.2-1. Although their appearance is rather forbidding, the structure can be understood by breaking the weights into groups. Since this is a clean-trained network, the bias and first input weights may be combined to yield a large effective bias of almost +30. It is convenient to divide the remaining weights into the two sets T-I2 through T-I9 and T-I10 through T-I32. The latter group is dominated by positive weights, of moderate strength; the longer the signal, the more this group will pull the cumulative sum toward positive values. For the longest signals, this contribution is significant, but not overwhelming.

The first group (T-I2 through T-I9) are mostly negative weights which process the most energetic portion of the signals. Ignoring temporarily the positive weight on T-I6, it is safe to say that this group as a whole will make a negative contribution to the cumulative sum. The slower the signal decays, the larger in magnitude is this contribution. The positive weight on T-I6 is not sufficient to prevent this. To see what effect this positive weight has, consider as a pair the inputs T-I5 and T-I6. For a signal which is steady or decreasing through these two inputs, it is clear that the contribution of this pair will be negative, due to the relative magnitudes of the weights. If more energy is present in input T-I6 than T-I5, the magnitude of the pair's contribution will be reduced. This is the case for some of the boomerang signals, such as B1P and B1W, in which the return of signal energy is increasing in this range.

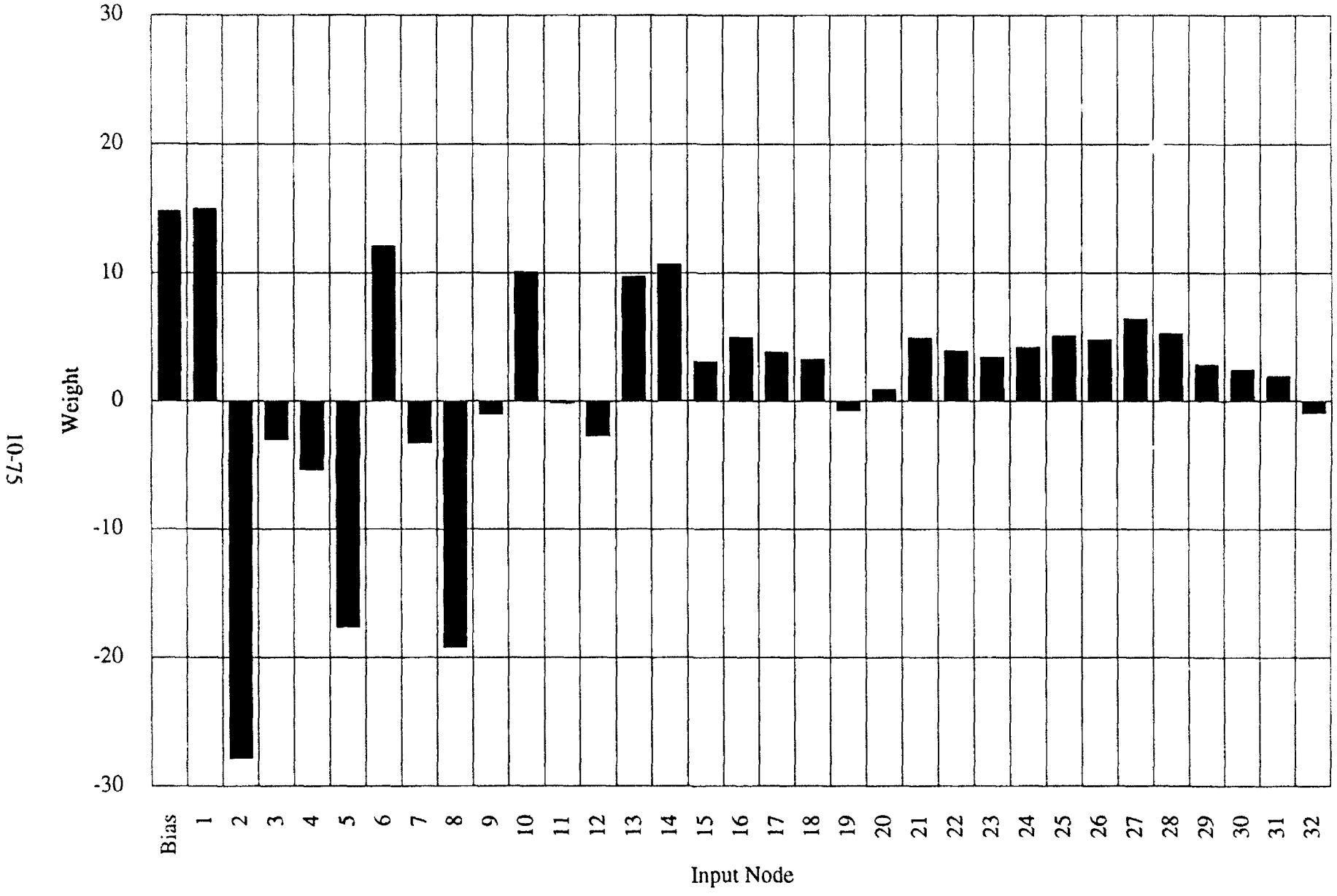


Figure 10.3.3.2.2-1 Weights on Input Layer to Hidden Node 2 Connections in Air4H(2)T

In summary, then, the node T-H2 starts with a very high positive effective bias. This is only reduced by energy in the early portion of the signal, so short signals tend to activate the node strongly. A slowly decaying signal can overcome the high bias in the first group of inputs, but may be pulled significantly back toward positive values by the last group of positive weights. The positive weight on T-I6 has relatively little effect on the activation of most signals, but may be a sensor of B1P and B1W signals, due to their unique shape.

The final issue is how the node's output responds to the various signals. A glance at this node's activations, shown in Figure 10.3.3.2.2-2, shows that they are virtually identical to those achieved by TN-H3 (see Figure 10.3.3.2.1-2) after the transfer function is applied (upper axes). Prior to the application of the transfer function (lower axes), it is clear that the two algorithms yield different results. Whereas the break between Brass 5% signals and the rest is the only definite division performed by hidden node TN-H3, T-H2 has in addition two clearly defined breaks which sort the signals further. As expected, only the longest signals (Brass 5%) were able to produce low activations of T-H2.

10.3.3.2.3 Comparison and Contrast of Hidden Nodes TN-H3 and T-H2

The inverse nature of the classifications performed by these two nodes is suggestive of how output layers use hidden nodes. The hidden nodes TN-H3 and T-H2 are almost perfectly negatively correlated, which implies that they sort signals into opposite orders. Yet, the output layers of the two networks do not use the nodes in opposite ways. This seems at first counterintuitive, but the activations shown in Figures 10.3.3.2.1-2 and 10.3.3.2.2-2 may help to clarify this point. Hidden node TN-H3 is used very heavily as a Brass 5% detector, because when it is activated, the applied signal is certainly of type B5M, B5P or B5W. On the other hand, when T-H2 is strongly activated, one can with certainty only make the statement that the applied signal is *not* a member of this class. This is a weaker statement because it means that the signal is from a Steel target, or 10% Thickness, or both. Which of these is the case is not accurately determined by the node T-H2, hence it is not used by the Material and Thickness nodes as much as TN-H3.

It was stated above that the general pattern of the input weights of TN-H3 (see Figure 10.3.3.2.1-1) is reminiscent of the input weights of another hidden node in the same network, TN-H2 (see Figure 10.3.2.2.1-1). The input weights of TN-H2 roughly resemble the inverse of TN-H3's weights. The correlation between TN-H2 and TN-H3 is only -0.44, however, so

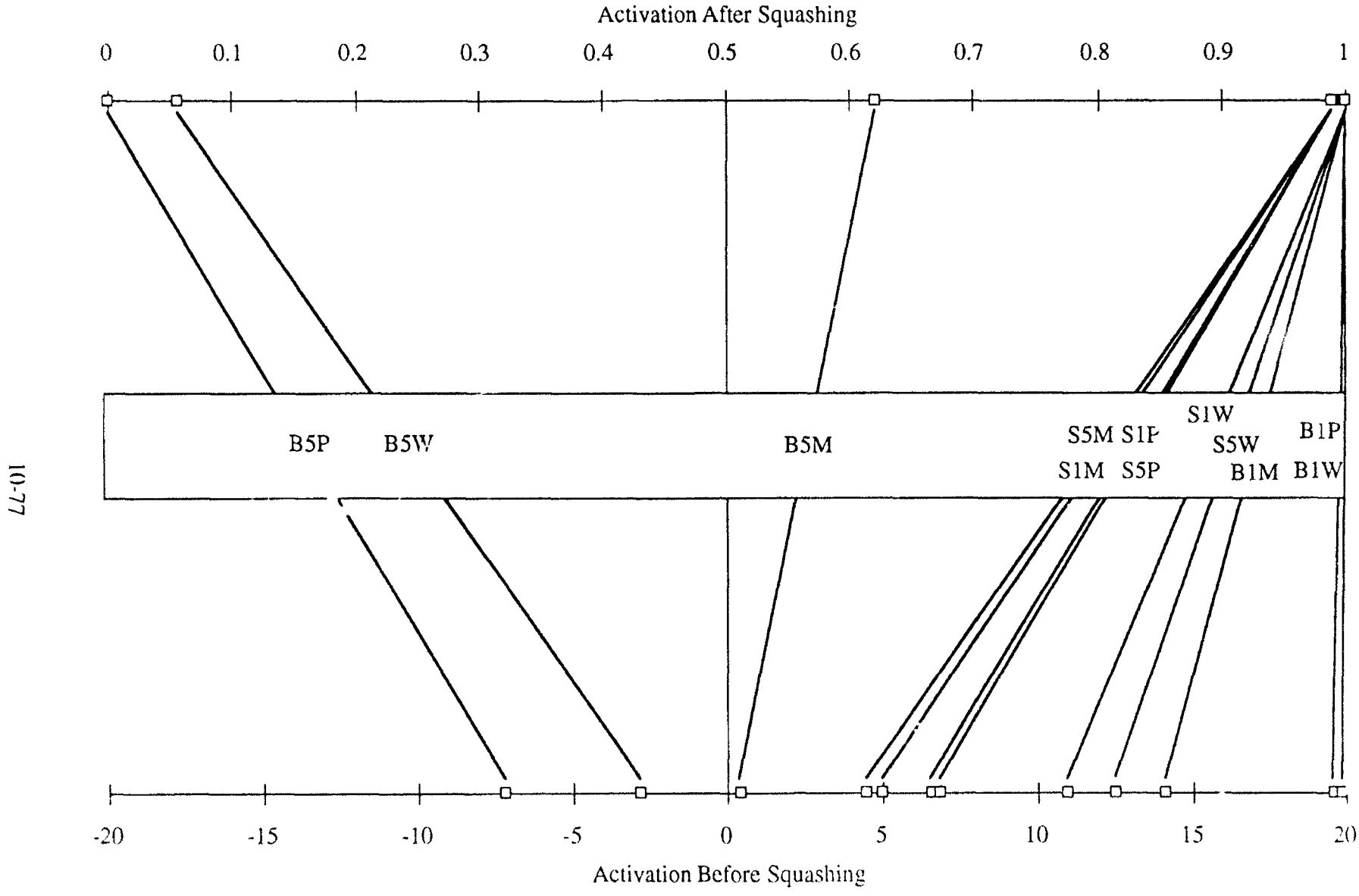


Figure 10.3.3.2.2-2 Hidden Node Air4H(2)T-H2 Activation for Instance Nine of Each Signal Class

despite the apparent similarity, the two nodes respond very differently to the various inputs. On closer examination of TN-H3, some differences may be observed in the input weights. In TN-H3, the input weights are noticeably "flatter" in the range TN-I2 through TN-I7. It is plausible that these more uniform weights respond to the total quantity of energy present in TN-I2 through TN-I7, while the tapered structure developed by TN-H2 is more sensitive to the decay characteristics of these inputs. Another difference between the two nodes is that the output of TN-H3 is more greatly affected by its negative bias than TN-H2 is affected by its positive bias. This will prove to be the critical difference between the nodes.

The activations of TN-H2 and TN-H3 (see the upper axes of Figures 10.3.2.2.1-2 and 10.3.3.2.1-2) produced by the various signals are quite different. The hidden node TN-H3 is only activated by three signal types, while TN-H2 is activated to varying degrees by a disjoint set of six signal types. Consider, however, the lower axes of these two graphs which show the cumulative sums of the nodes achieved by the signals. These graphs show that these nodes distribute the signals into exactly the opposite order. Moreover, the gaps between each signal and the next are proportionally almost the same for the two nodes. Prior to the application of the transfer function, then, the two nodes perform virtually the same (albeit inverted) calculation on the inputs. Of critical importance is how the signals are oriented relative to the origin. For example, if TN-H3 sorted the signals into the same order, but shifted their cumulative sums by approximately +13.25, the origin would be situated between the signals S1W and S5W. The final output of TN-H3 would then resemble very closely the inverse of TN-H2. This shift of the signals can be accomplished, merely by adding 13.25 to the input bias of TN-H3. The result of this transformation is shown in Figure 10.3.3.2.3-1. The differences between this graph and the inverse of Figure 10.3.2.2.1-2, which displays the activations of TN-H2, are very slight.

In conclusion, TN-H3 was found to be a very accurate detector of Brass 5% signals, while T-H2 is a rejector of this same signal type. This is consistent with the strong, negative correlation between the two. The algorithm developed by TN-H3 was simple, involved few inputs, and strongly resembled the inverse of that developed by TN-H2. The main qualitative difference between these latter two nodes lies in the relative strength of the bias weight. This subtle difference is sufficient to allow the nodes to respond very differently to the signal set.

The hidden node T-H2 used a very different algorithm, involving more of the inputs in a complex computation. The algorithm essentially balances the energy in the first nine inputs with the energy

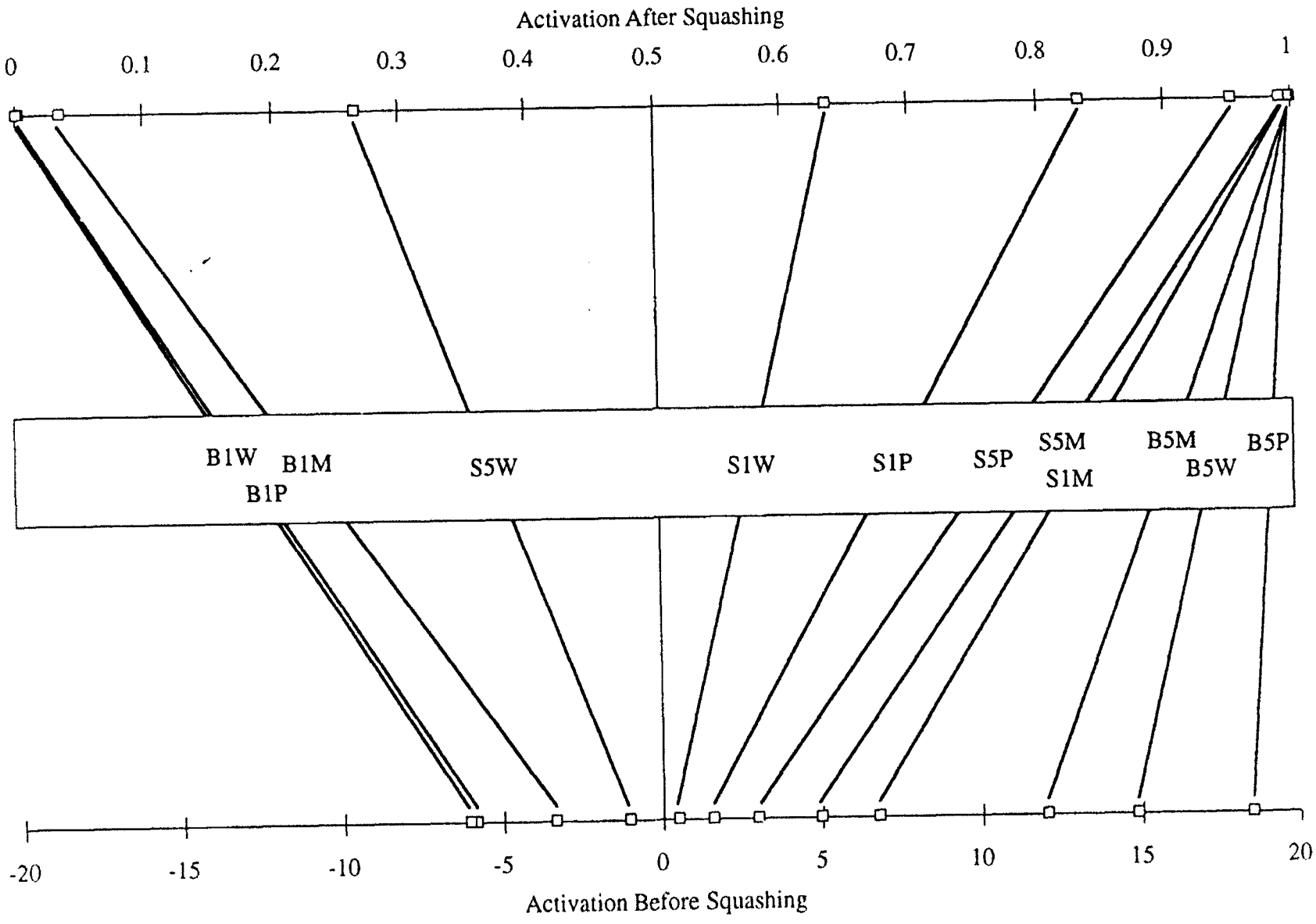


Figure 10.3.3.2.3-1 Hidden Node Air4H(2)TN-H3 Activation with Adjusted Bias for Instance Nine of Each Signal Class

in the rest of the inputs to categorize the signals. The algorithms developed by the network trained with noise focus on only the early portion of the signal, performing similarly by balancing the signal energy against a bias weight.

The presence of training noise helps the networks find a more general and robust solution. In the Air time domain networks, this took the form of simpler weight structures which rely most heavily on the early portion of the input (which contains the largest signal values). Different hidden nodes in the same network may perform almost redundant calculations and still provide different information to the output layer. This can occur because inverting the input weights and/or altering the bias can dramatically affect which signals activate the hidden node.

10.3.3.3 Comparison of TN-H3 and T-H2 to Best Second and N4 Third Dimensions

While the correlations found between the dimensions, Best second and N4 third, and the signals statistics indicated frequency domain relationships, these network nodes were able to produce correlations above chance levels with the dimensions using time domain signal input. The two scaling dimensions thus appear to have a dual time/frequency characteristic. In fact, the correlation between TN-H3 and the Best second dimension is due entirely to the high activation of TN-H3 by the Brass 5% signals vs. 0 activation for all other signals. We tend to reject the theory that the subjects applied the pure time domain strategy found at this node since the signals are more evenly distributed on the dimension than are the activations produced by the node.

The network devised a simple time domain strategy to perform its classification of the Brass 5% signals. This strategy consisted of rejecting all signals (using a large negative bias) which did not have significant energy relatively late in the signal. Only the Brass 5% signals passed this test. T-H2, trained without noise added to the signals, found a highly negatively correlated, but rather more complex, solution. While the listener is struck by the frequency domain differences between the Brass 5% signals and others, and frequency measures correlated best with this dimension, the network has demonstrated a time domain analog to this strategy which was not discovered through other means.

10.3.3.4 Analysis of Air4H(2)FN-H3

Hidden node 3 from Air4H(2)FN, FN-H3, is correlated with the Best 2nd dimension at -0.71 and with the N4 third dimension at 0.77. These relatively strong correlations are of particular interest because the signal measures correlated with these dimensions were both computed in the frequency domain, and this node used frequency domain input. The strategy on this node shed light on the arrangement of signals on the dimensions in question.

The dimensions were correlated with the mean and the modal frequencies of the signals from each class. The Best second dimension was positively correlated while the N4 third dimension was negatively correlated. Signals which had high mean or modal frequencies (which are themselves highly correlated) tend to be high on the Best 2nd dimension, and low on the N4 third dimension.

FN-H3 achieved its correlation with the dimensions by sorting the signals as shown in the activation chart in Figure 10.3.3.4-1. This shows that a group of Steel signals, S1M, S5M, S1P, and S5P, suppressed the node, while all other signals excited the node. There are no signals which produced moderate activation. This is the opposite of the means employed by the time domain nodes described previously to achieve high correlation with these dimensions. On both dimensions the 12 signal classes are distributed relatively evenly, with the Brass 5% group at one extreme and the Steel Metal Striker pair of signals (plus S5P in the case of Best second dimension) at the other extreme. The time domain hidden nodes differentiated the Brass 5% signals from all others, while this frequency domain node separates the Steel, Metal and Plastic Striker signals from all others.

These dimensions tend to separate the signals by Material, especially N4 third dimension. In keeping with this distinction, FN-H3 is used by the output layer as a detector of Brass signals. This makes sense since only two of the eight signals which activate this node are Steel. The node is also used as a detector of Wood Striker signals, and the four Wood Striker signals activate the node along with four other signals.

The weights from the input layer to FN-H3 are shown in Figure 10.3.3.4-2. Although they appear rather arbitrary, certain features are noticeable. The bias is large and positive. The largest positive weights are of lower frequency than the largest negative weights. A weighted average frequency computed on the positive weights would clearly be lower than that computed for the negative

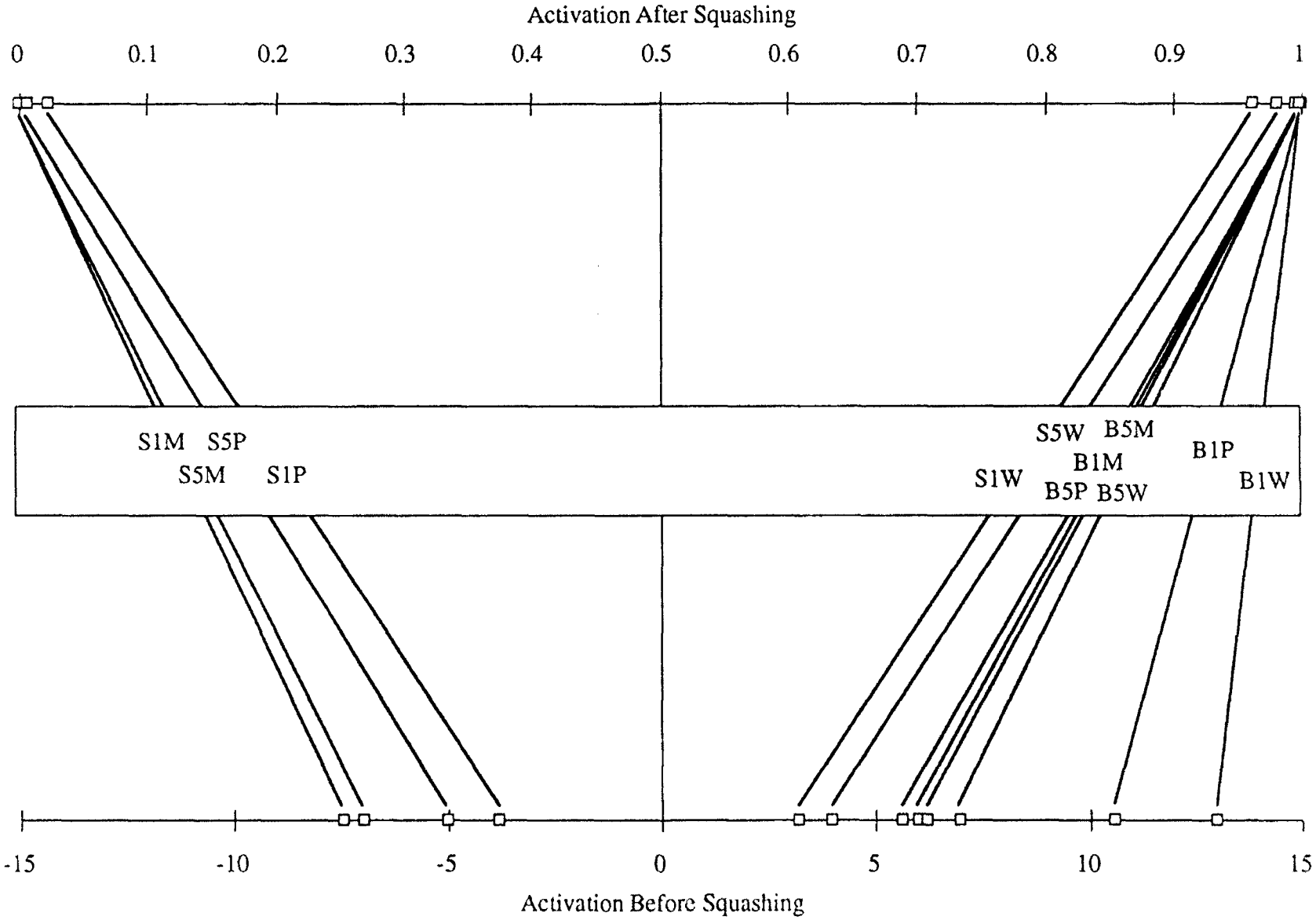


Figure 10.3.3.4-1 Hidden Node Air4H(2)FN-H3 Activation for Instance Nine of Each Signal Class

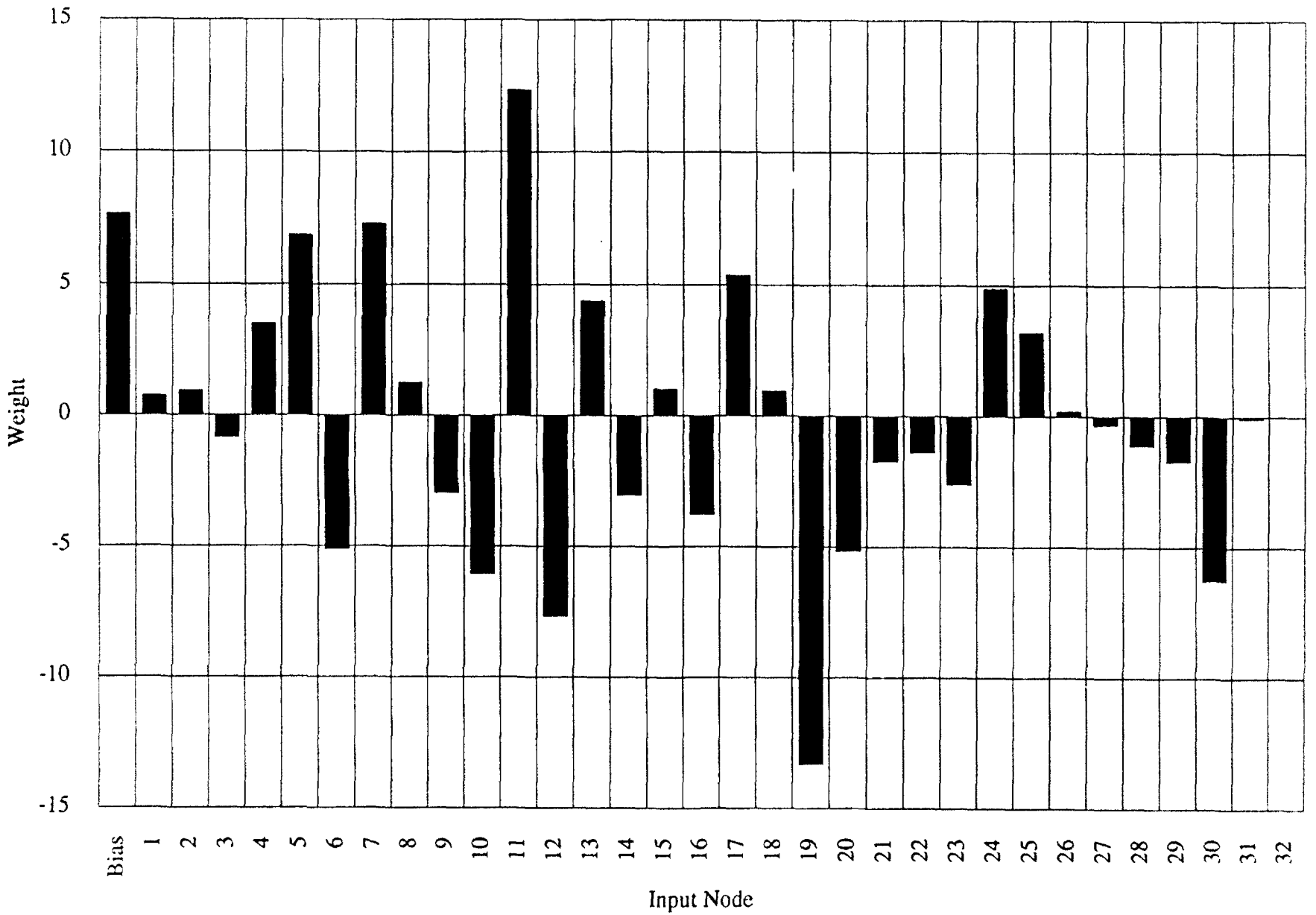


Figure 10.3.3.4-2 Weights on Input Layer to Hidden Node 3 Connections in Air4H(2)FN

weights. Thus, signals with their primary energy in lower frequency bins would tend to activate the node while signals with primary energy in higher frequency bins would tend to suppress the node. Since the node is negatively correlated with the Best second dimension, signals which suppress the node tend to be high on that dimension, and we saw above that the dimension is positively correlated with the mean and modal frequencies of the signals. Thus the weights considered by themselves tend to support the theory that this node applies a processing strategy similar to that found on the dimension, one based largely on average frequency content.

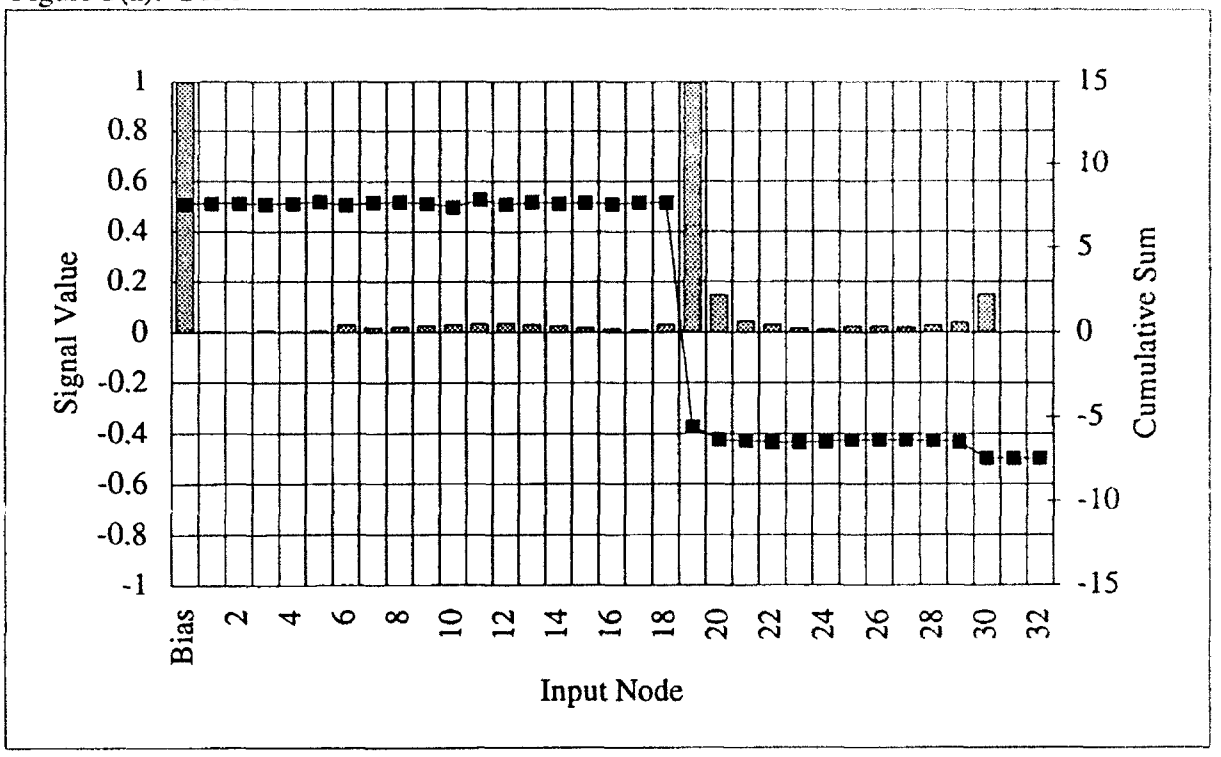
These observations were verified by analysis of the cumulative sums of the node when various signals were applied. All four of the signals which suppress this node do so exclusively using the large negative weight on bin 19. An example, S1M, is shown in Figure 10.3.3.4-3(a). With the exception of B1M (shown in Figure 10.3.3.4-3(b)), all other signals have their predominant energy, or peak energy, or both, at lower frequencies. Examples of these signals are shown in Figures 10.3.3.4-4(a) and 10.3.3.4-4(b).

This frequency domain node developed an activation strategy which produced results correlated with both of the human dimensions in question, and which in fact closely resembles the human processing strategy derived from the analysis of signal measures described earlier. In the case of FN-H3 a neural network node used largely the same processing strategy as that apparently used by the subjects to sort the signals into a highly related sequence. Meanwhile the time domain nodes that were correlated with the same dimensions found a strategy in the time domain which is related, in the sense of sorting the signals into another sequence highly related to the dimensions. These processing strategies in the time and frequency domains illustrate the potential of the networks to reinforce human strategies and to illuminate other potential strategies which might be employed.

10.3.4 Best Third Dimension

On the third dimension of the Best scaling solution the signals are divided by Material with one exception, S5M. At the high extreme are the three Brass 10% signals. At the low extreme are a group of Steel signals, S5W, S1P, and S1W. On the first dimension of the Best solution these six signals were grouped together to form one half of the dimension. Using the strategy of the third dimension, however, the subjects were highly sensitive to a difference between these groups. This strategy would also seem to be the primary means by which the three subjects as a whole achieved high performance discriminating the material parameter.

Figure 3(a): S1M



Signal Value — Cumulative Sum

Figure 3(b): B1M

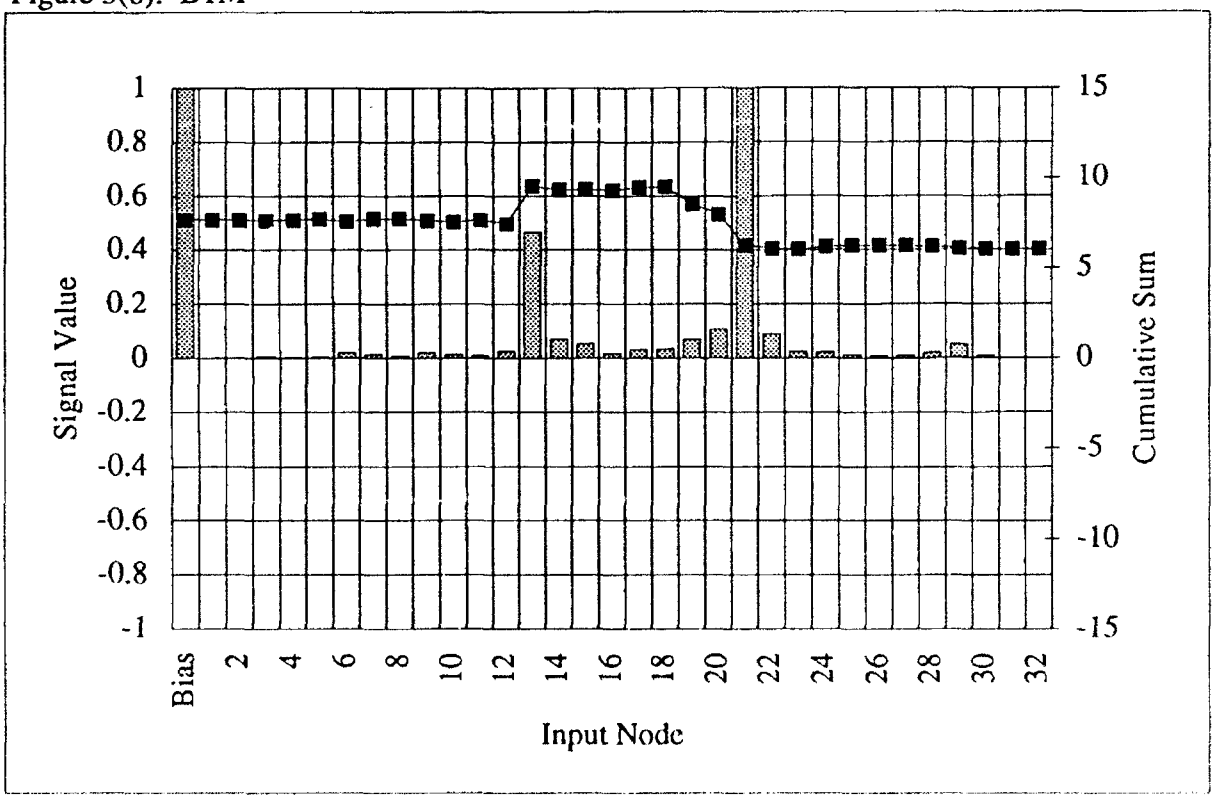
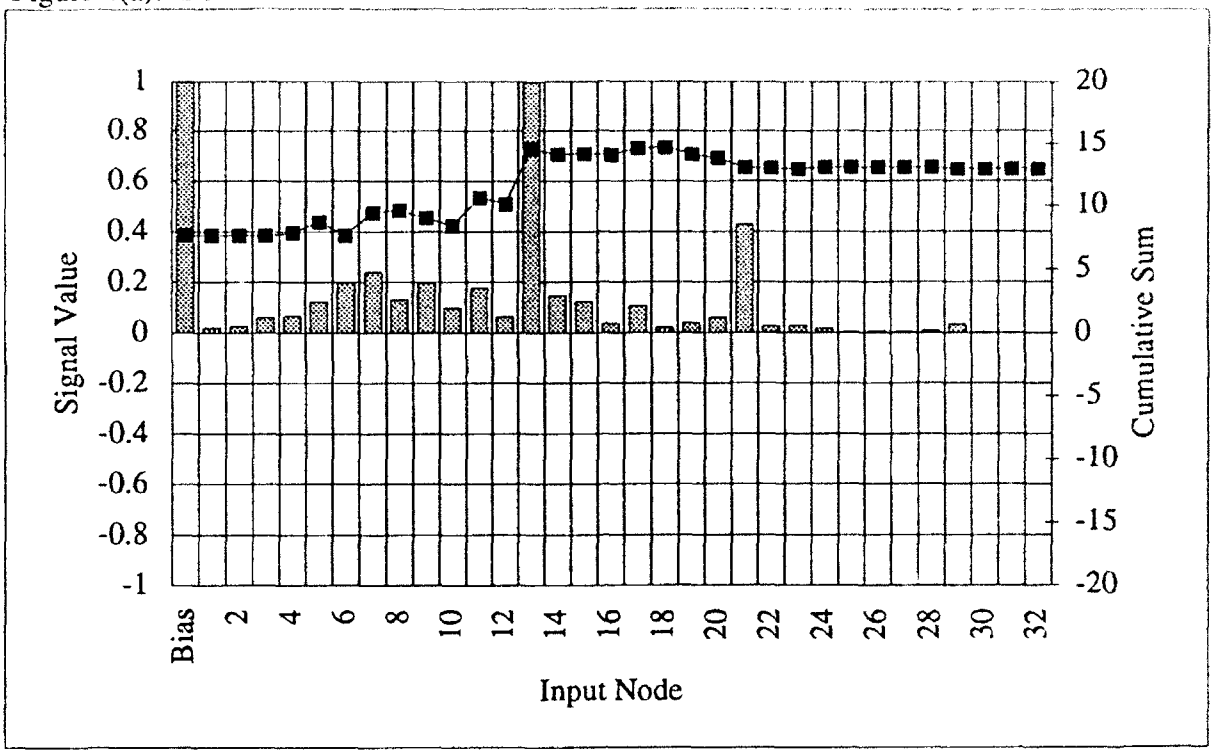


Figure 10.3.3.4-3 Cumulative Sum of Hidden Node Air4H(2)FN-H3
for Instance Nine of Classes S1M and B1M

Figure 4(a): B1W



Signal Value — Cumulative Sum

Figure 4(b): S1W

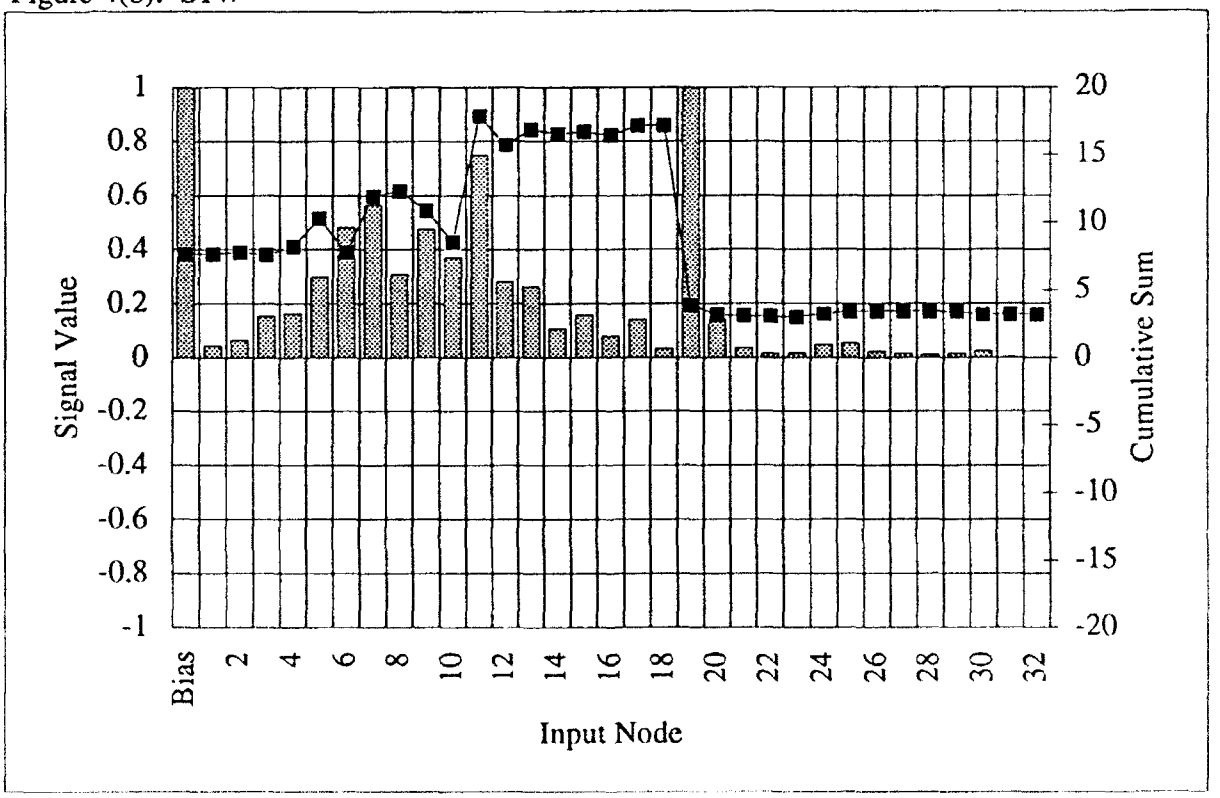


Figure 10.3.3.4-4 Cumulative Sum of Hidden Node Air4H(2)FN-H3
for Instance Nine of Classes B1W and S1W

10.3.4.1 Dimension Analysis

The only high correlation to this dimension is with low frequency slope, indicating a sharper cutoff of low frequencies in the signals higher on the dimension, such as the Brass 10% signals. This is consistent with the relationship between the Brass 10% signals and the third dimension of N4 at the second step of the regression.

When the highly damped Brass 10% signals are not considered, the dimension correlates quite well (-0.85, as shown in the plot of Figure 10.3.4.1-1) with the frequency of the most persistent sine wave in the curve fit solution. Listening verifies this relationship. The Brass 10% signals sound quite different from other signals in damping so quickly, and we speculate that the subjects processed this difference in duration along with the differences in frequency. They may have interpreted the lack of ring as a lack of high frequencies, which would place the Brass 10% signals high on this dimension.

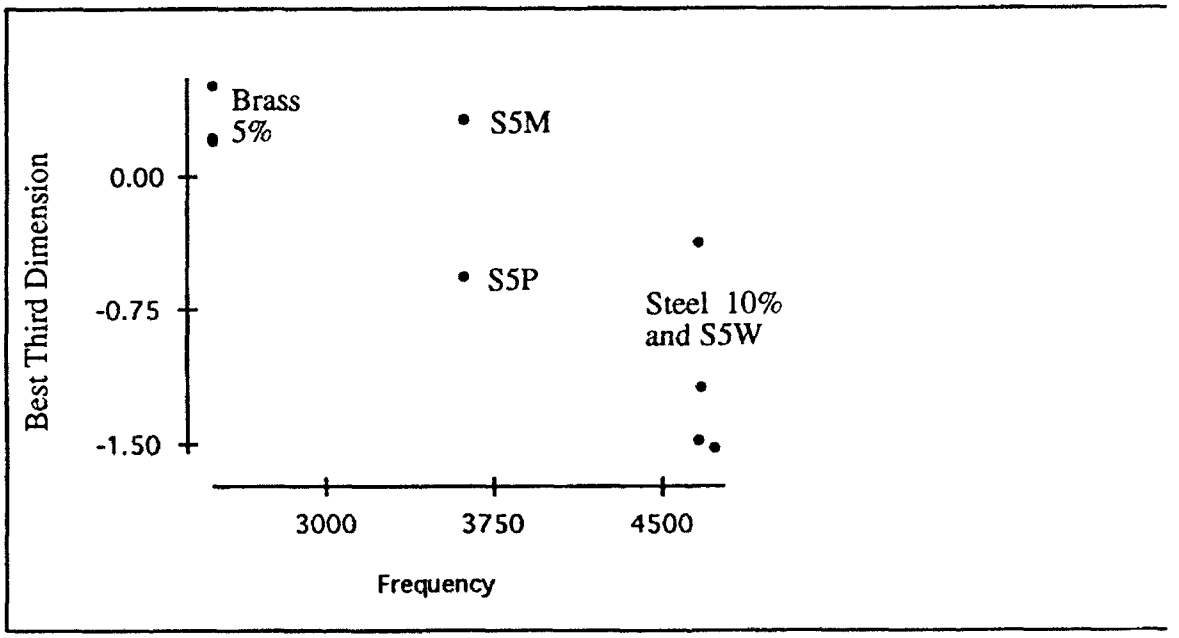


Figure 10.3.4.1-1 Frequency vs. Best Third Dimension Without the Brass 10% Signals

Low frequency slope is the leading candidate for inclusion in a regression, giving:

$R^2(adj) = 74.4\%$
Low Frequency Slope $p = 0.0001$

The inclusion of the amplitude of the most persistent sine wave is significant in accounting for the remaining variance:

$R^2(adj) = 82.4\%$
Low Frequency Slope $p \leq 0.0000$
Amplitude $p = 0.0424$

The presence of a time domain predictor is surprising given the correlation with a frequency domain measure as well as the impression made on the listener.

10.3.4.2 Analysis of Air4H(2)F-H1 and Air4H(2)FN-H1

Hidden nodes Air4H(2)F-H1 and Air4H(2)FN-H1, referred to for the rest of this section as F-H1 and FN-H1, are both correlated with the Best third dimension at 0.80 and with one another at 0.96. Both of these nodes are used by their respective networks to detect Steel signals.

10.3.4.2.1 F-H1 Analysis

Since the network trained without the addition of noise to its inputs classified the material parameter perfectly, and F-H1 is the only means of doing so, we may safely assume that the node was activated by Steel signals and suppressed by Brass signals. This was verified by the activation graph shown as Figure 10.3.4.2.1-1. This also served to explain the high negative correlation with the Best third dimension, which tended to sort the signals by Material with Steel signals low on the dimension. The high correlation between the activations of the two nodes indicated that they produced quite similar outputs.

The input weights of F-H1, shown in Figure 10.3.4.2.1-2, appeared rather complex. There were several frequency bins by which a signal could be detected or rejected. However, the various classes of signal interacted with these weights in a limited number of ways. The Brass 10% signals were rejected due to high energy in bins 13 and 21 (see Figure 10.3.4.2.1-3(a)). Each Brass 5% signal was rejected due to its energy in bins 10 and 13 (see Figure 10.3.4.2.1-3(b)).

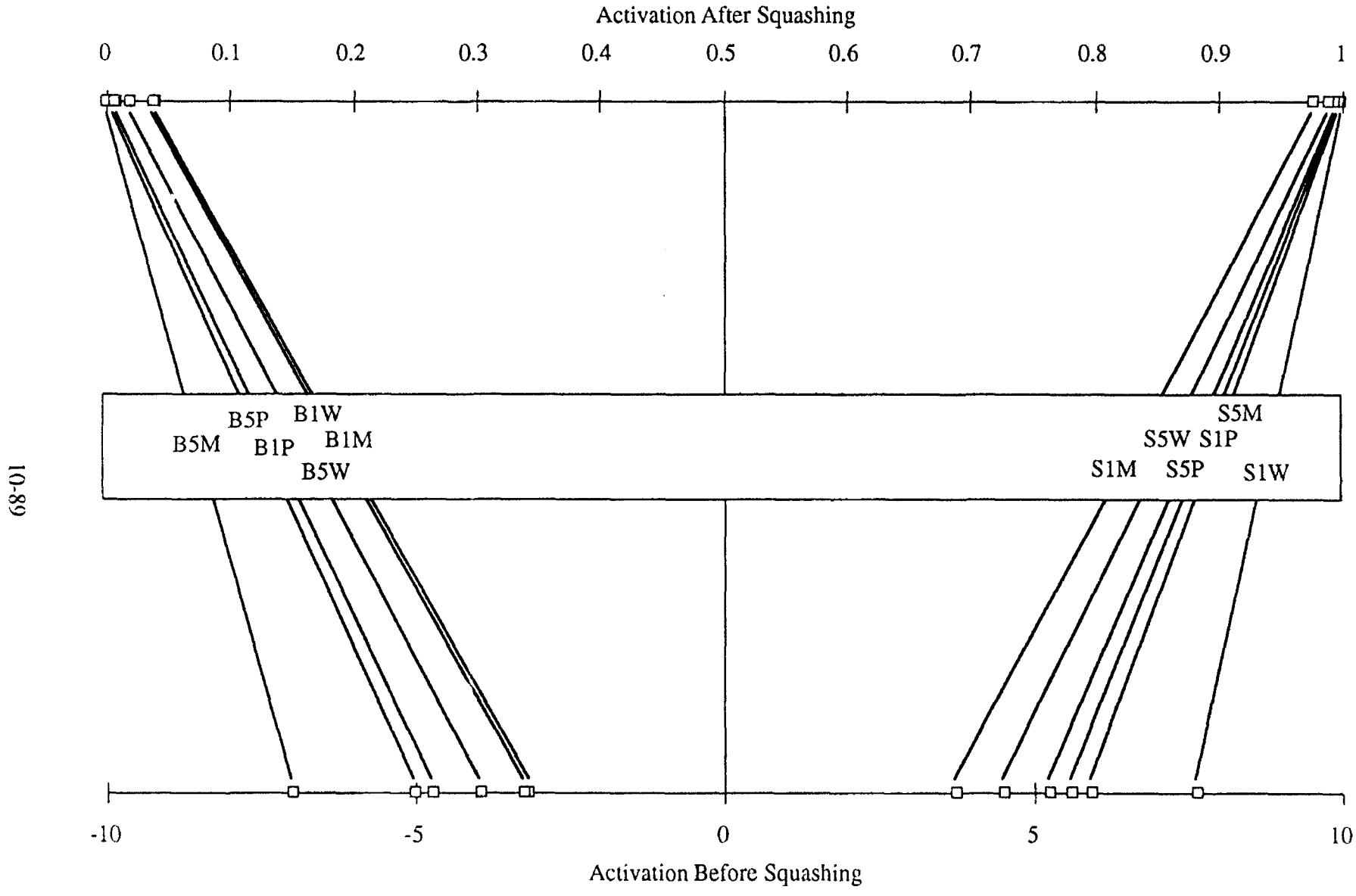


Figure 10.3.4.2.1-1 Hidden Node Air4H(2)F-H1 Activation for Instance Nine of Each Signal Class

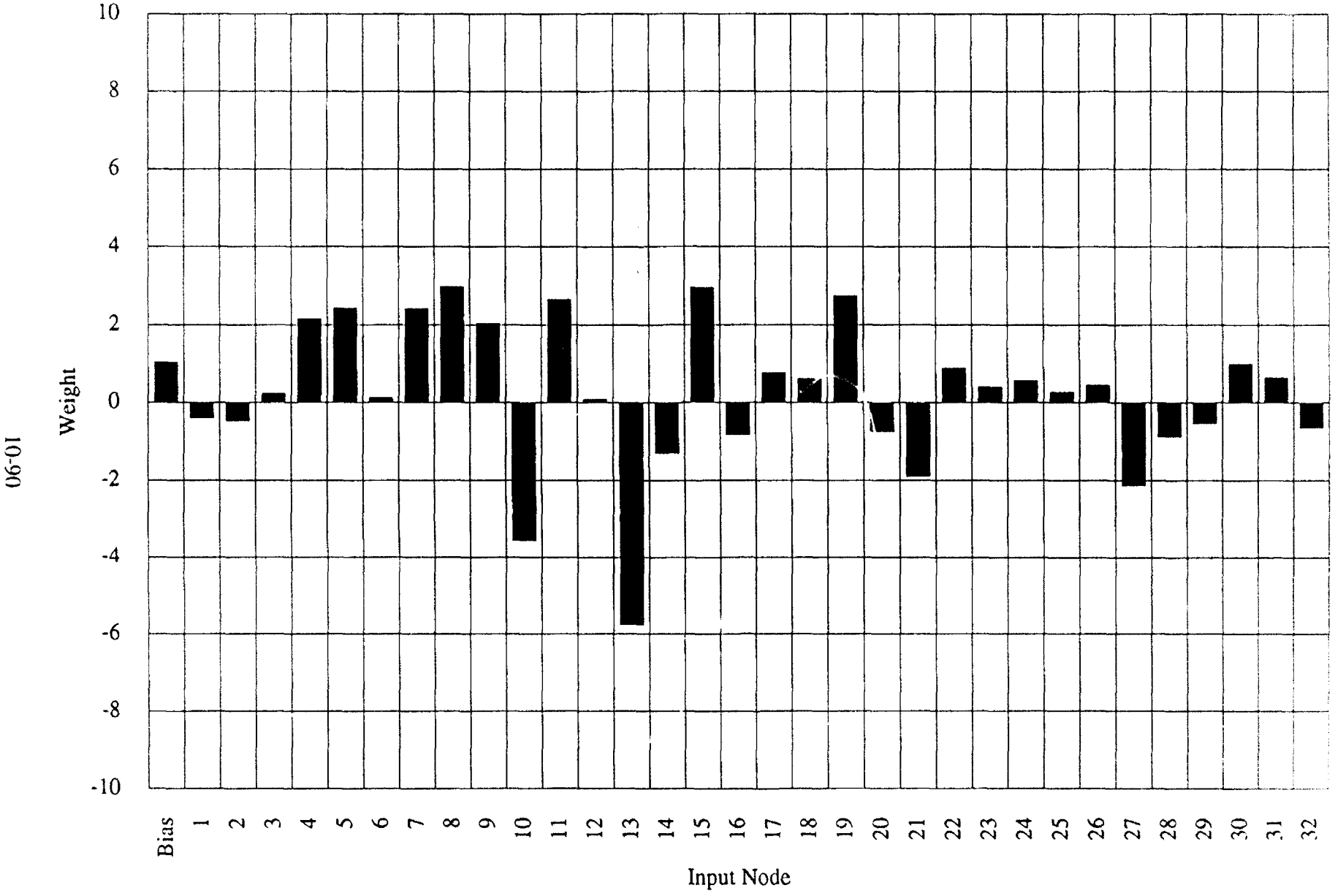


Figure 10.3.4.2.1-2 Weights on Input Layer to Hidden Node 1 Connections in Air4H(2)F

Figure 3(a): B1M

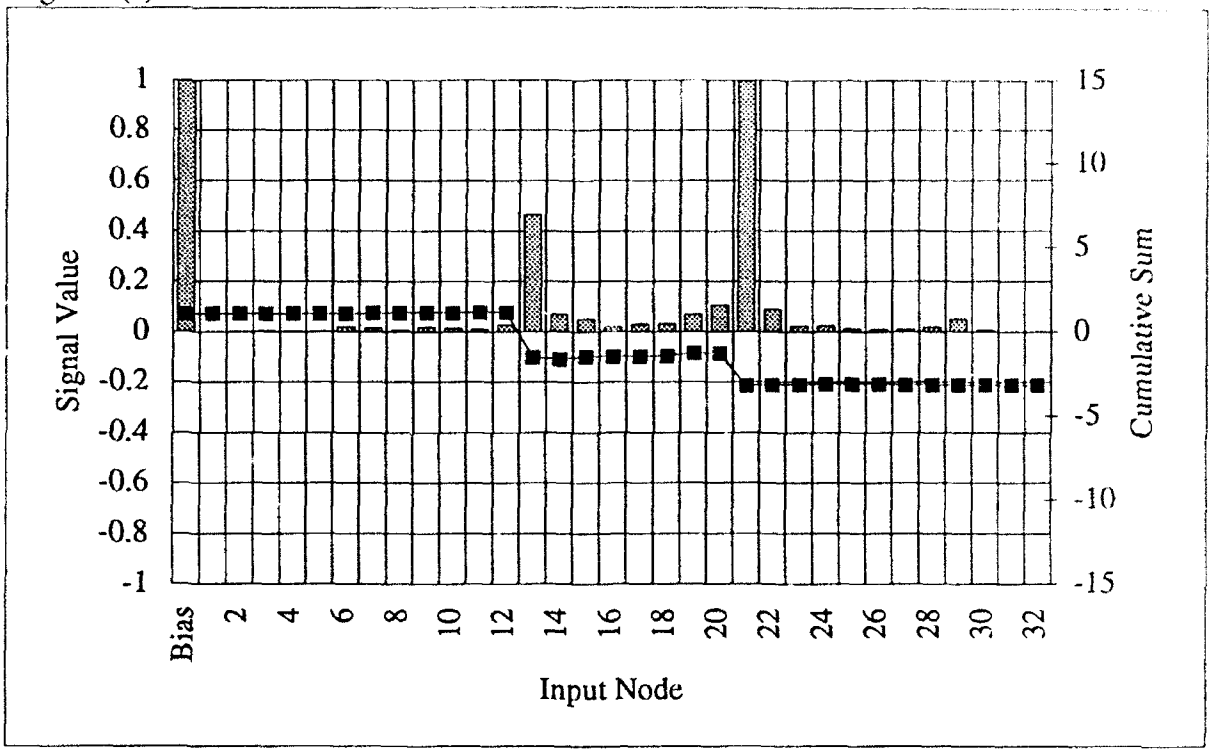


Figure 3(b): B5M

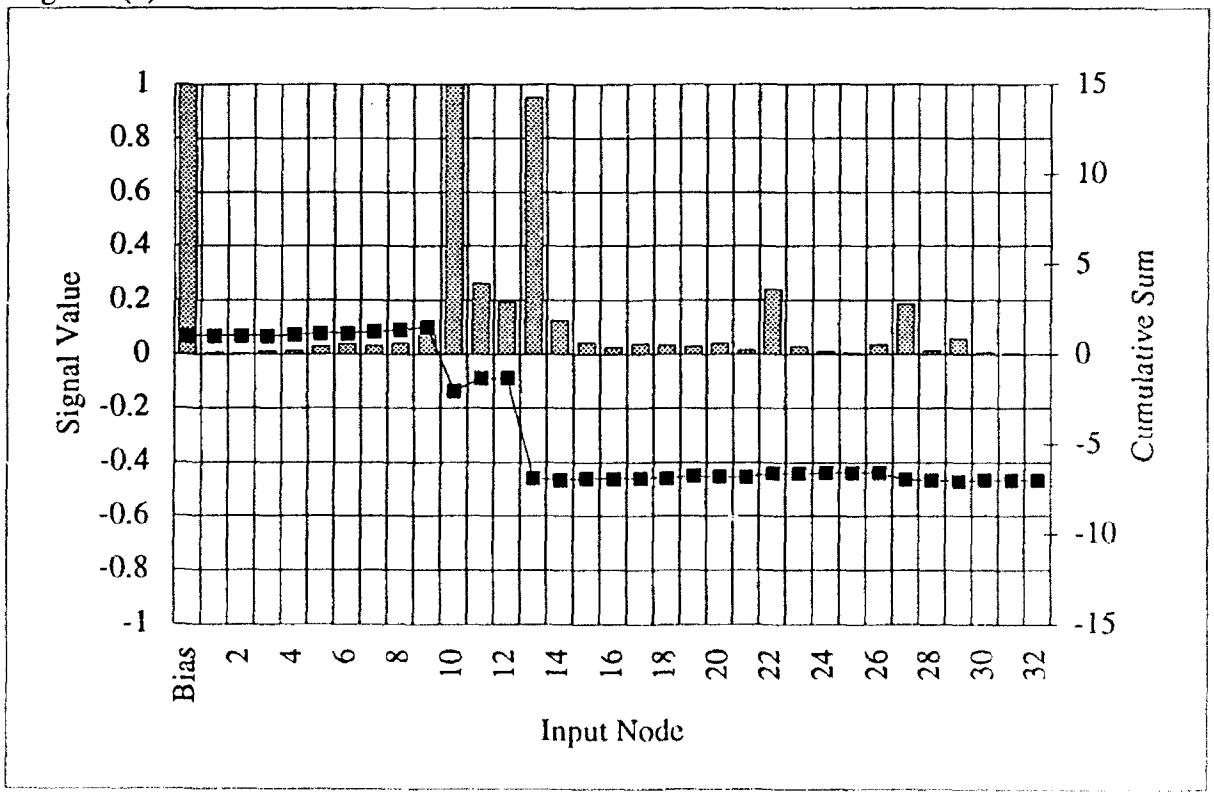


Figure 10.3.4.2.1-3 Cumulative Sum of Hidden Node Air4H(2)F-H1 for Instance Nine of Classes B1M and B5M

The Steel 5% signals, along with S1M, are detected because they have high energy in bins 15 and 19 (see Figure 10.3.4.2.1-4(a)). S1P and S1W activate the node using a broad range of high energy in bins 3 - 13, and with a peak in bin 19 (see Figure 10.3.4.2.1-4(b)).

10.3.4.2.2 Analysis of Air4H(2)FN-H1

Hidden node Air4H(2)FN-H1 is quite similar to Air4H(2)F-H1. The input weights of FN-H1 are shown in Figure 10.3.4.2.2-1. The signs and relative magnitudes of the weights are almost all the same as those of F-H1. There was decreased relative emphasis on bins 15 and 19, although these weights are still high enough to play the same roles as in F-H1, and increased relative emphasis on bin 17. The sorting order is quite similar, as seen in Figure 10.3.4.2.2-2, with the exception of the Brass Wood Striker signals, which receive moderate activation instead of none. The higher bias of FN-H1 helped to account for this.

10.3.4.2.3 Comparison of F-H1 and FN-H1 to Best Third Dimension

In summary, the weights in the range I4 to I11 serve to detect S1P and S1W, which have broad high energy in this range. Other high weights are tuned to particular subclasses, including I10. These peaks are highly reminiscent of the relationship described above (high negative correlation) between the frequency of the most persistent sine wave component of the signals, and the dimension excluding the Brass 10% signals. The Brass 10% signals have no ringing portion, and may be processed as a special case by the subjects. The remaining Brass signals (5%) peak in energy at bins 10 or 13, while all of the Steel signals peak at the higher frequency bins 15 or 19. The Steel signals appeared to have generally higher frequency components than the Brass 5% signals, and by this characteristic the node produced high activations for the Steel. signals.

10.3.5 N4 Second Dimension

This is the only dimension with a partial breakdown of the signals into groups by Striker. Probably not coincidentally, this subject was the highest performer on the striker parameter (59% vs 46% and 43%). The Metal striker signals are separated from the rest and placed low on this dimension. Whatever signal feature the subject used to distinguish the Metal signals apparently did not succeed with the other Strikers, as they are mixed. However, the remaining signals are divided

Figure 4(a): S5P

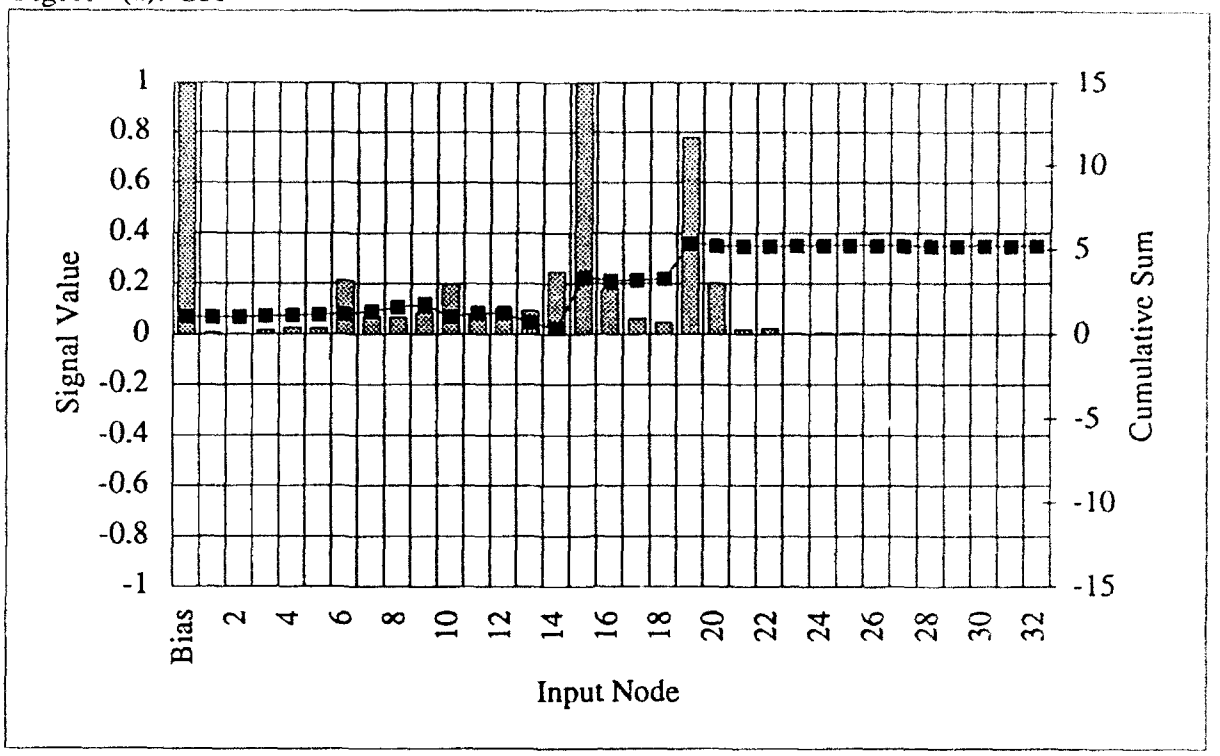


Figure 4(b): S1P

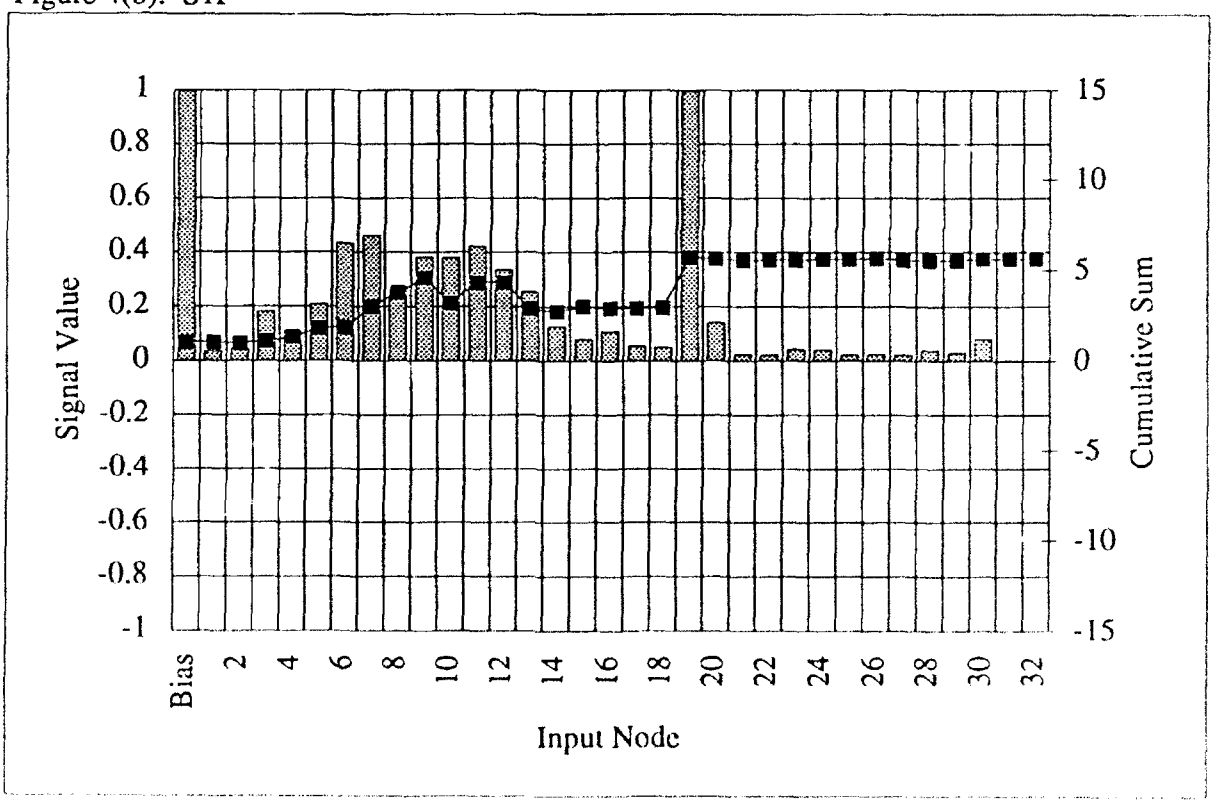


Figure 10.3.4.2.1-4 Cumulative Sum of Hidden Node Air4H(2)F-H1 for Instance Nine of Classes S5P and S1P

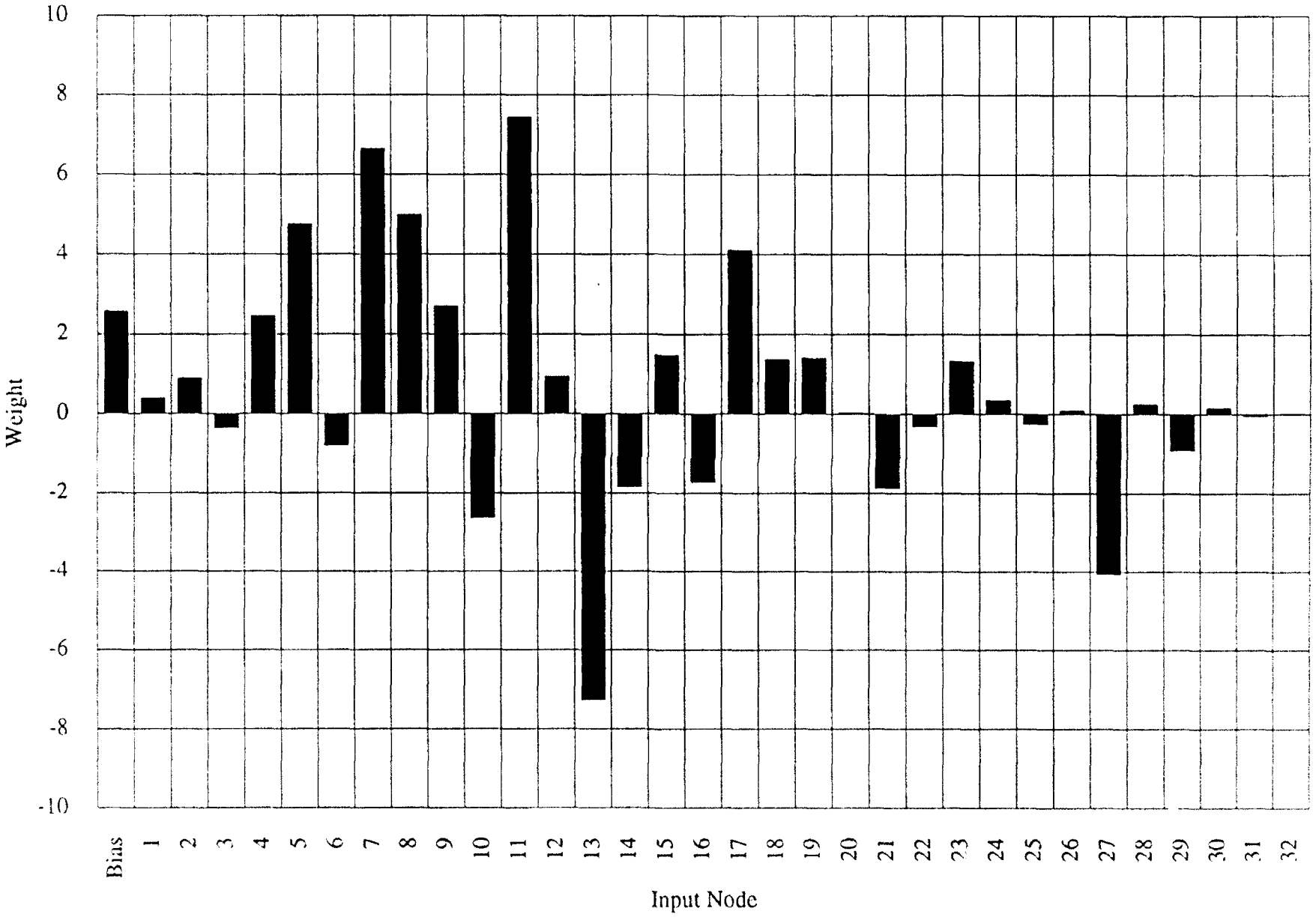


Figure 10.3.4.2.2-1 Weights on Input Layer to Hidden Node 1 Connections in Air4H(2)FN

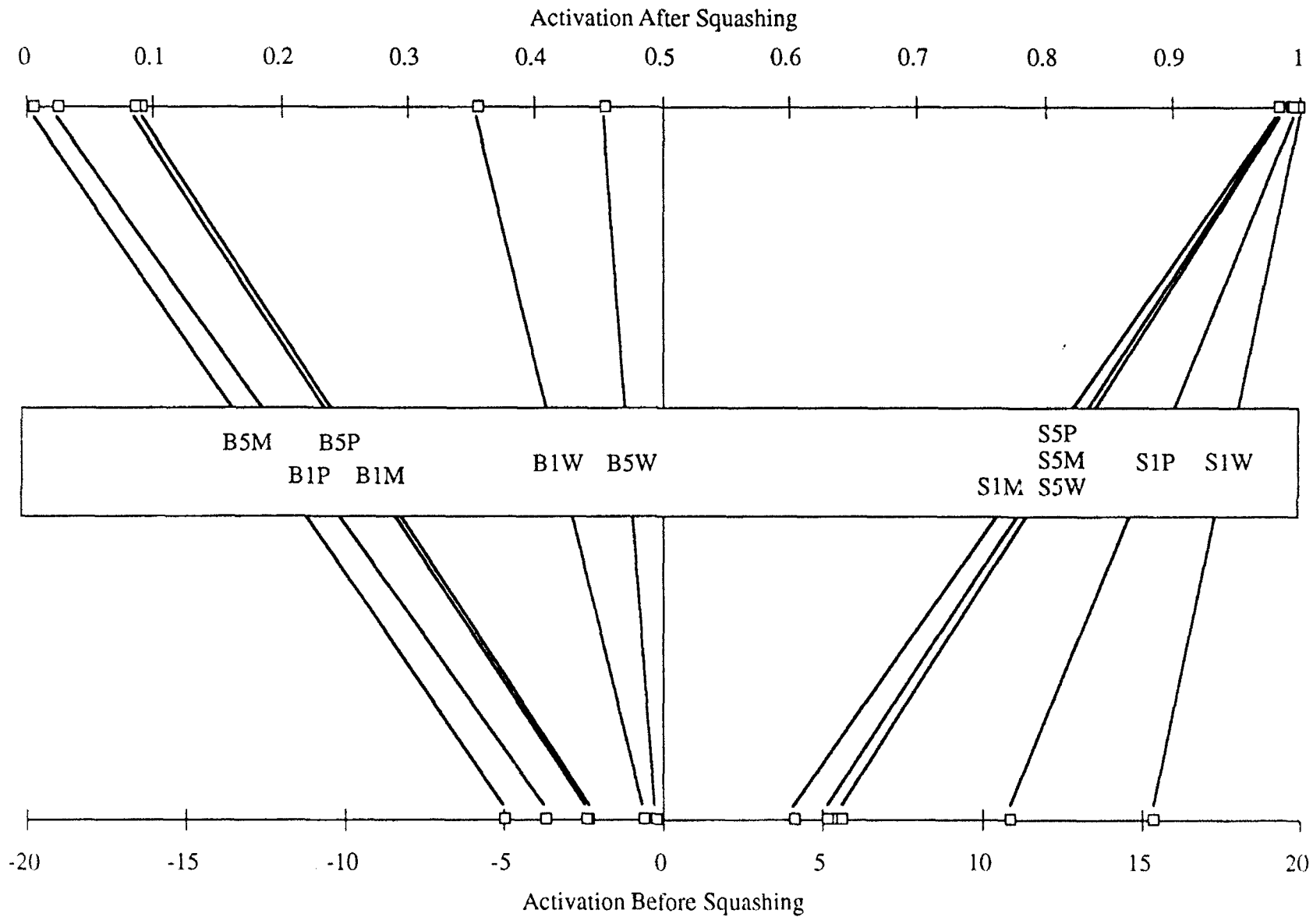


Figure 10.3.4.2.2-2 Hidden Node Air4H(2)FN-H1 Activation for Instance Nine of Each Signal Class

by Material, with the four remaining Brass signals the next higher group, followed by the four remaining Steel signals highest on the dimension. The four Brass signals are also grouped by Thickness, in keeping with their groupings on the other two dimensions of N4. Presumably the signal feature associated with this dimension sorted the signals into groups of Metal striker, remaining Brass, then remaining Steel.

10.3.5.1 Dimension Analysis

The only significant correlation between this dimension and a statistical measure is with the initial amplitude of the most persistent sine wave of the curve fit solution. This high negative correlation (-0.83) can be recognized when listening to the signals. The four Metal signals are lowest on this dimension and have the highest initial amplitude. Higher on the dimension are signals which may ring as long as any other signal, but which start from a lower amplitude. This relationship is captured best by the amplitude measure, which applies only to a sine wave at a single frequency, found by the curve fit algorithm. Correlation with the "decay amplitude," which accounts for all energy in the signal, was lower at 0.60. This may indicate that the subject was not sensitive on this dimension to any distinct impact sound or to frequencies other than the longest-lasting. Listening to this dimension suggested that the subject was focusing attention on the onset of the ringing portion of the sound, the magnitude of which was captured reasonably well by the amplitude measure as described above.

Amplitude is the best single predictor of the dimension. This is so because the amplitude measure rates the Metal signals and the Brass 5% signals, as one group, higher than the remaining signals. This is shown in Figure 10.3.5.1-1.

It would appear that this is a good approximation of the technique used by the subjects on this dimension. Using amplitude as the predictor yields:

$R^2(adj) = 65.7\%$
Amplitude $p = 0.0008$

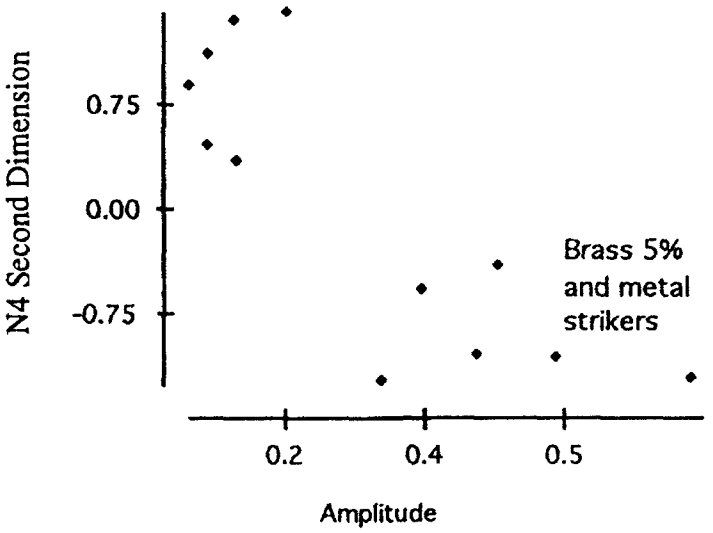


Figure 10.3.5.1-1 Amplitude vs. Second Dimension

10.3.5.2 Analysis of Air4H(2)F-H4 and Air4H(2)FN-H4

Hidden nodes Air4H(2)F-H4 and Air4H(2)FN-H4 (henceforth referred to as F-H4 and FN-H4) were correlated with the N4 second dimension at -0.80 and -0.81 respectively. The nodes are very highly correlated with one another (0.99), and achieved that correlation with almost identical weights. Because the weights are essentially the same, only F-H4 is discussed below.

No time domain hidden nodes correlated significantly with the dimension. Although the most highly correlated signal measure was the initial amplitude of the most persistent sine wave found in the curve fit solution, the envelope of the signal which was presented to the time domain networks did not include information at such a fine level. If the initial amplitude were indeed a good description of the signal processing that the subject was using on this dimension, the time domain neural networks had no way of using the same strategy.

Both F-H4 and FN-H4 were used by their respective output layers as detectors of Metal strikers, as seen in Figures 10.3.2.3.1-1 and 10.3.2 3.2-1. The output layers were expecting these nodes to isolate the Metal striker signals in much the same manner as did the dimension. Indeed, the

response of the F-H4 to all of the signal classes (seen in Figure 10.3.5.2-1) shows that the node produced high activation only for the Metal signals. The node did not arrange the remaining signals as the dimension did, but performed the single critical task of detecting the Metal signals.

The strategy by which F-H4 detected the Metal signals is embedded in the hidden weights shown in Figure 10.3.5.2-2. The first logical group of weights is I1 - I11, which are negative except for the very small weight on I1. Since these weights would serve to suppress the node, and the node is suppressed by Plastic and Wood signals, it is reasonable to look for high energy in Plastic and Wood signals in this region (and to expect Metal signals to have little energy in this frequency band). Input weights I12 - I32 are generally positive, with two exceptions (I14 and I16). This group of weights has more variability in weight values than the first group. A few weights are very strong, indicating greater selectivity among the frequency bins when identifying Metal signals.

The bias on F-H4 is approximately 2, which tends to activate the node, but not strongly compared to many of the weights. The bias is easily overcome by the product of the weights in the band I1 - I11 for Plastic and Wood signals which, unlike Metal signals, contain significant amounts of energy at these frequencies. Plastic and Metal striker signals thereby suppress the node. These signals do not take advantage of the negative weights on I14 and I16.

Three of the four Metal striker signals rely on the bias for detection. That is, they have so little energy in the portion of the frequency band which is negatively weighted that the modest bias remains the major component of the sum on the node. These signals are illustrated by the hidden node response to B1M, shown in Figure 10.3.5.2-3(a). The Brass 5% Metal signal class is the only exception. As shown in Figure 10.3.5.2-3(b), it has enough energy in the band I1 - I11, particularly I10 and I11, to suppress the node by interacting with the weights that normally process Plastic and Metal signals. To overcome this, the network developed positive weights at I12, I13, I22, and I27. These are present only to produce high activation for B5M.

10.3.5.3 Comparison of F-H4 and FN-H4 to N4 Second Dimension

F-H4 and FN-H4 developed a simple method of identifying Plastic and Wood striker signals and differentiating them from Metal striker signals. For the single case of a Metal signal which meets the criteria set forth by the nodes for Plastic and Wood signals, the nodes developed a special case.

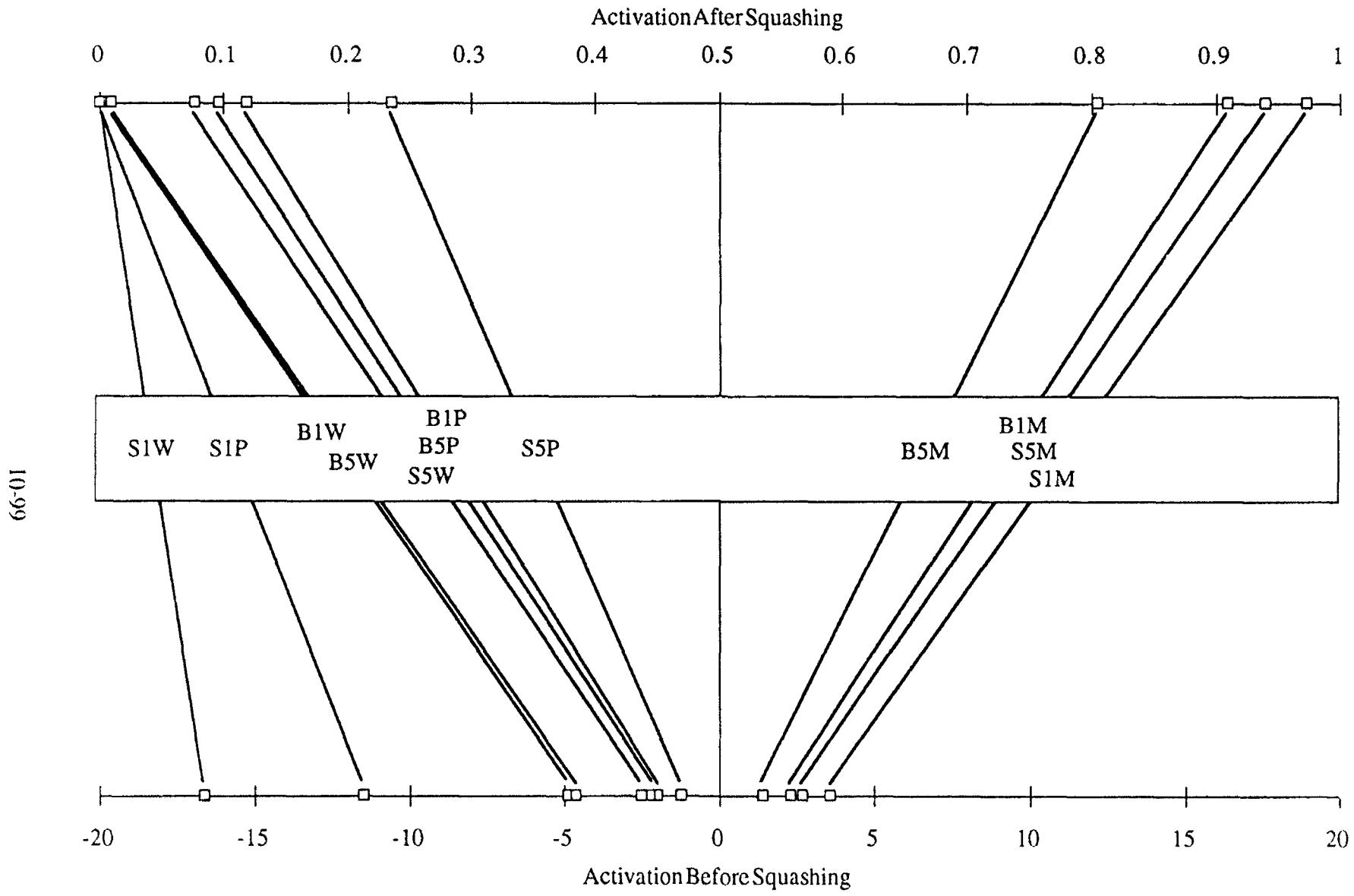


Figure 10.3.5.2-1 Hidden Node Air4H(2)F-H4 Activation for Instance Nine of Each Signal Class

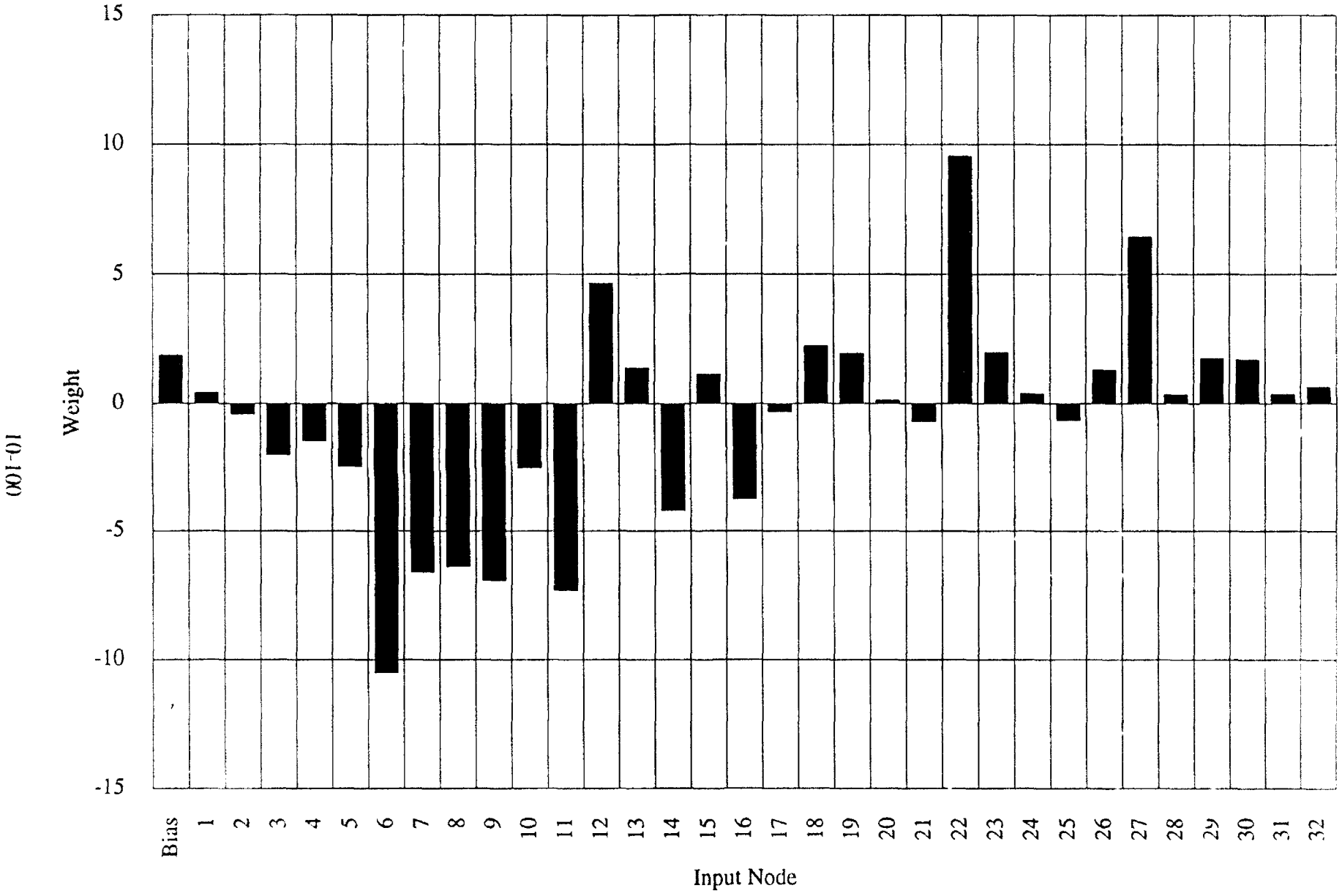


Figure 10.3.5.2-2 Weights on Input Layer to Hidden Node 4 Connections in Air4H(2)F

Figure 3(a): B1M

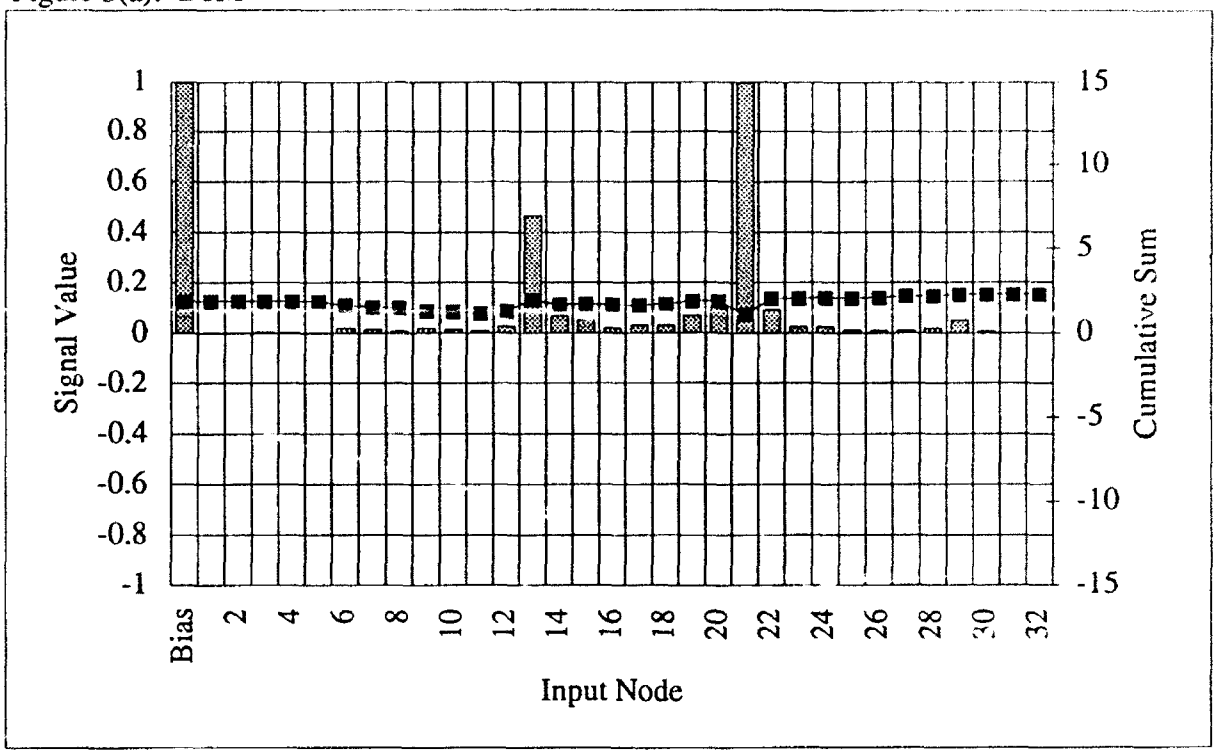


Figure 3(b): B5M

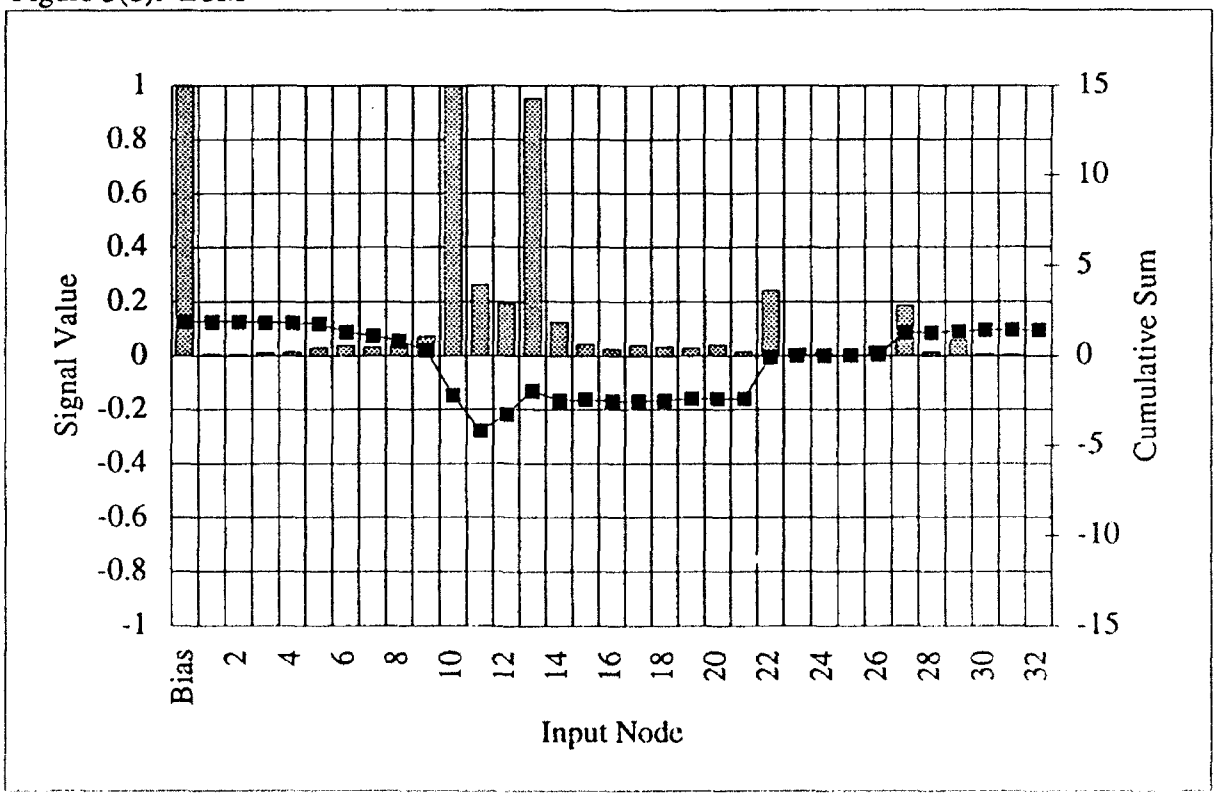


Figure 10.3.5.2-3 Cumulative Sum of Hidden Node Air4H(2)F-H4
for Instance Nine of Classes B1M and B5M

Both nodes found a very general solution to the problem of differentiating the Metal signals from Plastic and Wood signals based on frequency domain input, and in the single case in which the general solution would not work the nodes developed a special case to handle that signal.

While no network developed a time domain node with activations highly correlated to this dimension, the form of signal input may have prevented this. At the least, the envelope form of time domain input prevented these networks from developing the strategy described above and based on the initial amplitudes of particular frequencies. The best four hidden node time domain neural network achieved 84.4% performance on Striker however, indicating that the envelope form of input did contain information about the Strikers. The subjects, on the other hand, did have both time and frequency information to use. There would appear to be some processing in both domains by the subjects, given the relationship between the dimension and the initial amplitude of specific frequencies in the signals. Given only the frequency information with which to differentiate the signals, networks were able to perform essentially the same processing as this dimension. Although the subjects used frequency domain information as well as time domain, to a large extent this node found a purely frequency domain method for making the same distinctions among the signals that the subjects did using substantial time domain information.

10.4 SUMMARY

The comparison of subject dimensions, signal measures, and network nodes illuminated the comparative processing strategies of subjects and networks. The signal measures were the initial means of modeling the dimensions created by scaling confusion data from the classification experiment. Lacking a direct means of observing subject acoustic processing, the signal measures were a means of examining the dimensions and implying processing strategies at a useful level of detail. These measures were not always easy to develop or to apply. The model of a dimension depended on the choice of appropriate signal measures, forcing the researcher to make assumptions about the likely means by which the subjects approached the classification tasks. Nevertheless, the models derived from signal measures are reasonably accurate predictors of the placement of signals on the dimensions and appear to the listener to describe legitimate processing strategies for the given signals.

Analysis of the hidden nodes which were correlated to the human dimensions proved feasible and very informative. In fact the development of correlated hidden nodes emerged as a practical means

of investigating human processing strategies on the dimensions. comparison of network and human processing revealed both strong similarities in processing methods under proper signal conditions as well as differences which could be exploited by human listeners. Experience with these analysis methods suggests several insights.

- Scaling dimensions capture essential elements of the subjects' processing strategies.
- The dimensions can be modeled at a useful level with readily available signal measures, with limitations on the depth of the models stemming from the relative lack of complexity of the signal measures.
- Neural network strategies to accomplish the same task as subjects may be essentially identical if the signal input provides the same information that the subjects used. This is particularly evident if noise is added to time domain signals.
- Networks will derive related strategies if signal input is in a different form than that used by subjects.
- Networks may be used directly to explain human processing when networks nodes are correlated to human dimensions and signal input is in an appropriate form.

11.0 DISCUSSION

The project gave insight into several areas of acoustic processing by people and by neural networks. These are discussed here in terms of the relative performances of the two subject groups, and of the subjects vs. the neural networks. The effects of applying low signal-to-noise ratio signals to the network's performance and processing strategies are discussed. The relationships among human dimensions, which are assumed to represent human processing strategies, and the network hidden nodes are discussed. Finally, logical extensions of the work are mentioned.

11.1 CLASSIFICATION PERFORMANCE

The human subjects were confronted with very difficult tasks in attempting to classify the underwater sounds. Many of these discriminations were too difficult for any subject to make, as indicated by the several performance levels at or near chance. Under these conditions, any possible differences between the subject groups were generally masked. Nevertheless, the Navy subject group performed significantly higher than the student group on one aspect of the Bottom signals, suggesting a difference in capabilities which a more reasonable task might illuminate.

The Air signals were created to provide a classification task of reasonable difficulty. These signals proved much easier for both subject groups to classify, while still providing the confusions needed by the scaling algorithm. Most subjects classified each of the three parameters above chance levels. When faced with this task of moderate difficulty, performance differences between the two subject groups emerged. Navy subjects as a whole were significantly better than the student group on several aspects of the Air signal classifications tasks. Student subjects were never significantly better than the Navy group.

Properly configured and trained neural networks performed much more highly than the human subjects. Much of this difference is due to the signal transformations necessary for the networks (necessary to increase performance and meet size and processing time restrictions). For instance, one can see the differences between signals in the frequency domain form used as network input, and the networks also found these differences. The subjects however probably could not always hear these differences, particularly in the underwater signals. In addition, neural nets are notorious

for finding artifactual differences between input classes. This tendency was part of the motivation for adding noise to the input signals.

Networks which used frequency domain input had higher performance than networks which used time domain input. Both kinds of input are highly processed from the original state of the signals, and the information content may not be comparable due either to that processing or to inherent limitations of the domains. Certainly the time domain inputs lose much information when they are enveloped and downsampled, as do the signal spectrums when they are averaged. Both techniques tend to reduce the quantity of artifactual information available to the networks as they reduce the signal information to manageable levels. Frequency domain signals may nonetheless contain more information useful to the networks than time domain signals.

The human subject were better classifiers of the Air signals than of Bottom or Free-field signals. Networks, on the other hand, performed slightly better on the underwater sounds than on the Air signals. Within the Air classification task, both subjects and networks found Striker to be the most difficult parameter to classify. The immediate information about Striker was short-lived, leaving the classifier to infer information about Striker from the ensuing signal.

11.2 *EFFECTS OF ADDING NOISE TO SIGNALS*

Both time and frequency domain neural nets were tested using low signal-to-noise Air signals as input. This noise was added at the input layer of the networks. The networks proved moderately robust to noise, with performance falling off steadily but not precipitously as noise was added to the signals. It is assumed that these noise levels would have proven quite difficult for human subjects. When noise was added during the training of networks, and the same tests on noisy signals were made, the resulting networks were significantly more robust to noisy test signals. While some networks did not improve or actually did worse, the large majority of networks increased their classification performances over a wide range of input noise.

the hidden nodes of time domain networks trained with noisy signals typically departed from those of networks trained without noise. When comparing two nodes which produced highly correlated activations for the various signals classes, one node trained with and one without noise on the inputs, the node trained with noise typically had a radically different weight pattern. This weight pattern implemented a much simpler processing strategy than did the weight pattern of the node

trained without noise. This phenomenon did not extend to the frequency domain networks, in which noise did not appear to have a significant effect on most hidden nodes (although classification performance was usually higher for the network trained with noise).

11.3 HIDDEN NODES AND HUMAN DIMENSIONS

The multidimensional scaling technique combined with the modeling of the scaling dimensions to produce concise explanations of human processing. Of course we cannot observe the subjects' processing directly, and so must rely on inferences based on confusion data. We cannot check these models against the physiological processes of the subjects, and so we assume they are a reasonable explanation of subject processing along the dimensions. These analysis techniques yielded explanations of each relevant dimension. These models were generally in the time domain or the frequency domain with little overlap, and certainly inform of only part of the processing of the subjects. Nevertheless they provide good explanations of the arrangement of the signals on the dimensions.

Neural networks attempting to classify the signals develop hidden nodes which often sort the signals into very similar patterns to those of the dimensions. In fact, each dimension was correlated to multiple hidden nodes. These hidden nodes were often of both time and frequency domain, even when correlated to a dimension which was modeled only in one of the domains. When a particular hidden node was trained in the same domain as the model of the correlated dimension, in most cases that node employed the same strategy as that of the model of the dimension. Neural network hidden nodes often developed the same strategy in classifying the signals as did the human subjects.

In the time domain, the nodes with the highest level of similarity to the dimension model were trained using noisy inputs. These nodes employed virtually the same strategies are their human counterparts, at least at the level of the models of the human dimensions. When a correlated node had been trained without noisy inputs, it employed a more complex but clearly related strategy. Nodes trained with frequency domain data usually showed no difference in strategies between those nodes trained with and without noise. The strategies, however, bore close resemblance to those of the correlated dimensions.

Some dimensions appeared to reflect strategies of the subjects which were applied only in one domain. Network nodes from the other domain were nevertheless able to sort the signal classes quite similarly. Such a capability might be suggestive of strategies that the subjects could employ, particularly subjects who have not learned to extract all possible information from a signal.

Experience with the Integrator Gateway Network suggests that these networks can also process the signals in a manner similar to that of subjects. When the confusion data from a Bottom IGN was scaled, the first two dimensions were similar to those of the subjects. The first dimension of the IGN was very highly correlated with both of the first scaling dimensions of the subjects, while the second dimension of the IGN was moderately correlated with the two second dimension from the subject results. This network had the same difficulties with the signals in the Bottom set that the subjects experienced.

11.4 EXTENSIONS OF THE RESEARCH

Within the current signal set, several logical extensions of the research may make sense. Network techniques have not been exhausted. One might be interested in the weight structure of networks trained to produce the same output as that of a subject attempting to classify the signals. The input form of the signal would be critical, but a network which successfully mimics human performance may provide insight into how the person achieved that performance. The differences between high and low performers could be investigated in this manner, as well as differences between various signal input transforms.

Explanations of the dimensions analyzed here might also be forthcoming from the weight structures of networks trained to replicate the dimensions. Again the complexities of signal input transforms would be critical to the information gained from the weights.

The human data has also not been fully tapped. Dimensions were derived only from top Navy performers. Differences in processing strategies between high and low performers, and Navy and student subjects, may be of interest. Finally, the techniques of the research should be applied to data more in keeping with the Navy subjects' typical acoustic processing tasks. These arbitrary signals do not reflect sonar technicians' typical environment nor level of difficulty.

NOTES

1. F.W. Young and D.F. Harris, *SPSS Base System User's Guide*, Ed. M.J. Norusis (SPSS, Inc.,Chicago, 1990), Chap. 25.

2. F.W. Young and R.M. Hamer, *Multi-dimensional Scaling: History, Theory & Applications*, (Erlbaum, Hillsdale, NJ, 1979).

3. Young and Harris, *SPSS User's Guide*, 428-451.

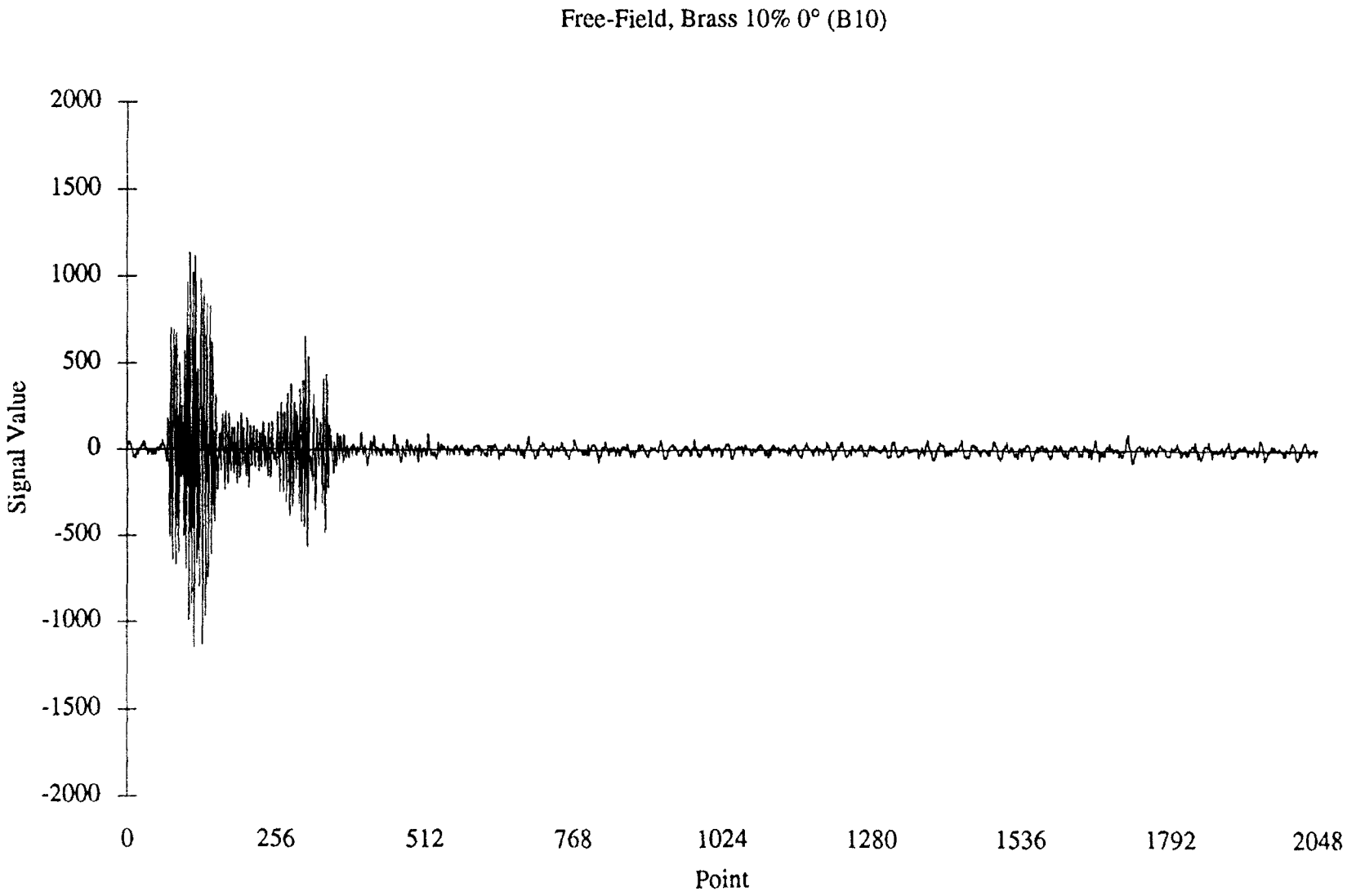
4. P.W.B. Moore, et.al, "Recognizing Successive Dolphin Echoes With an Integrator Gateway Network," Neural Networks, **4**, No. 6, 701-709 (1991).

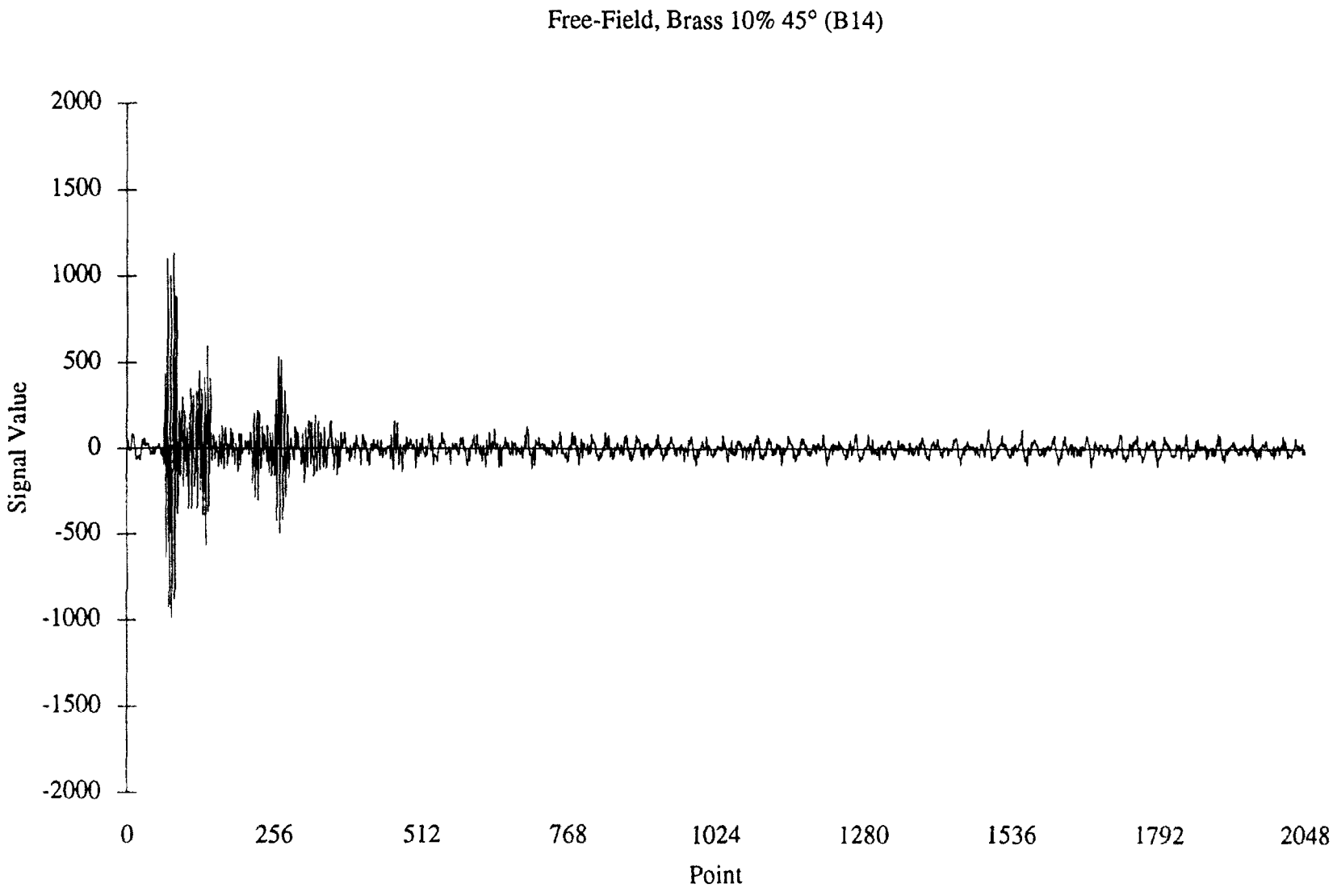
5. W.H. Press, et. al., *Numerical Recipes in Pascal*, (Cambridge University Press, Cambridge, 1989).

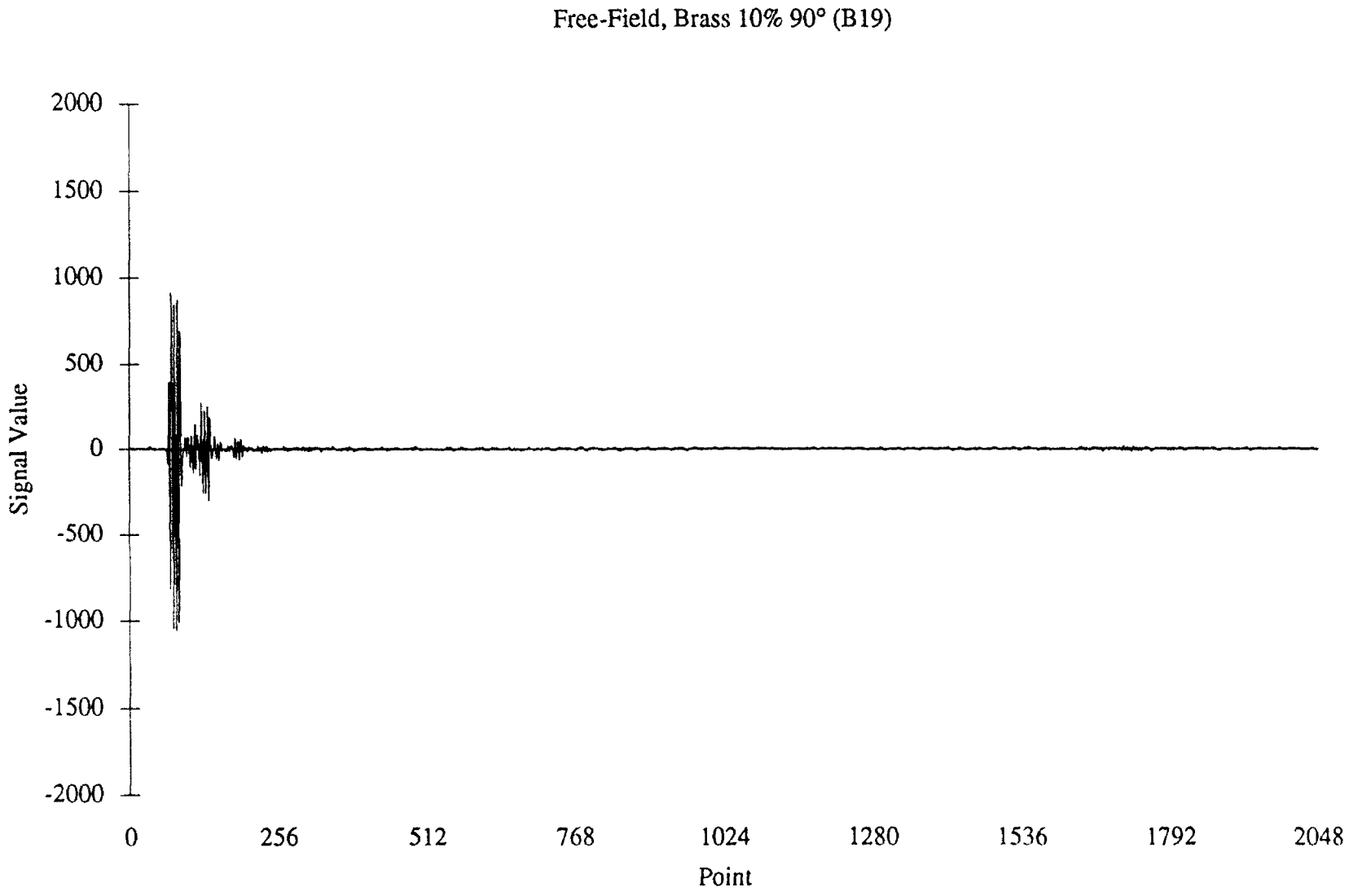
APPENDIX A

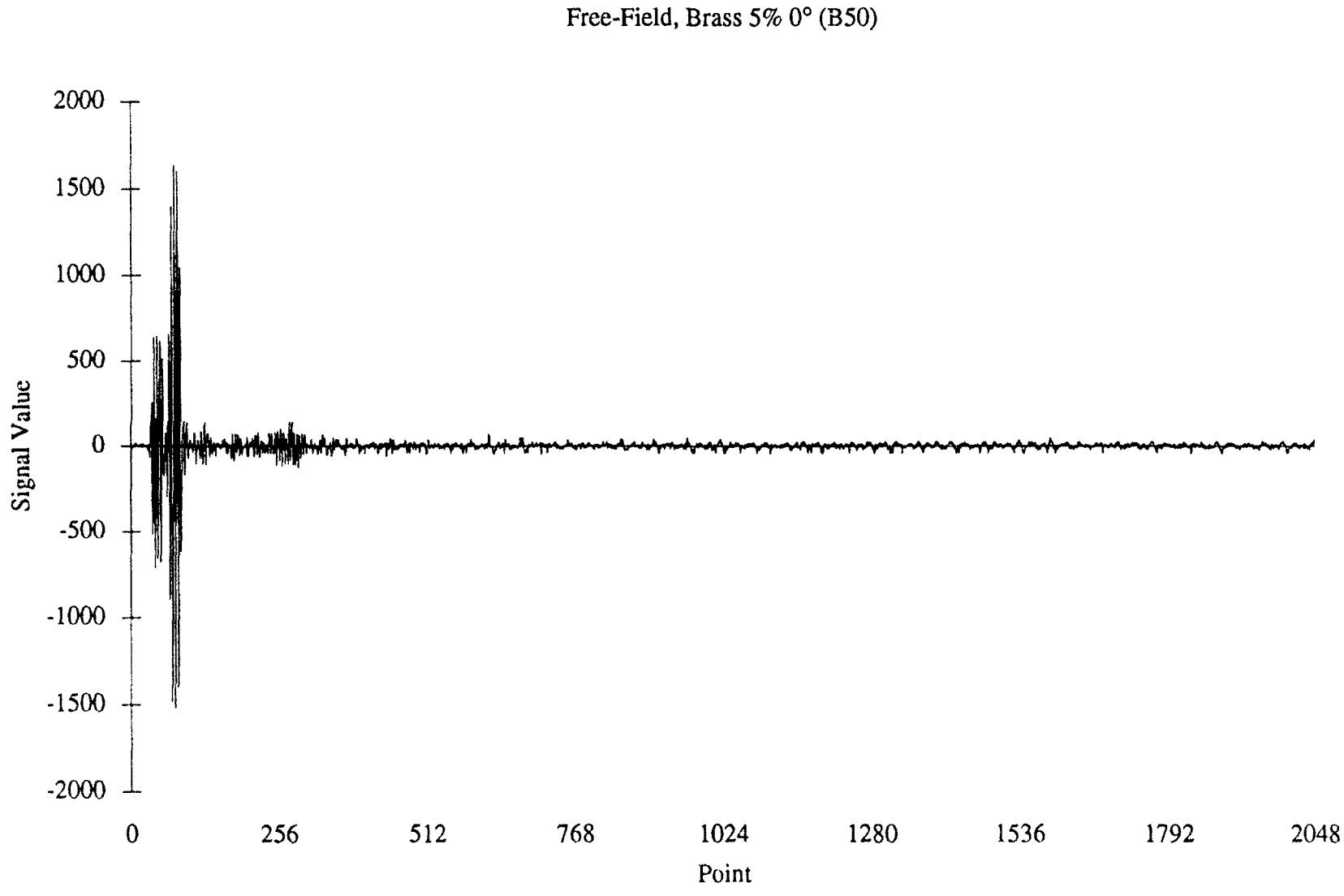
FREE-FIELD SIGNALS

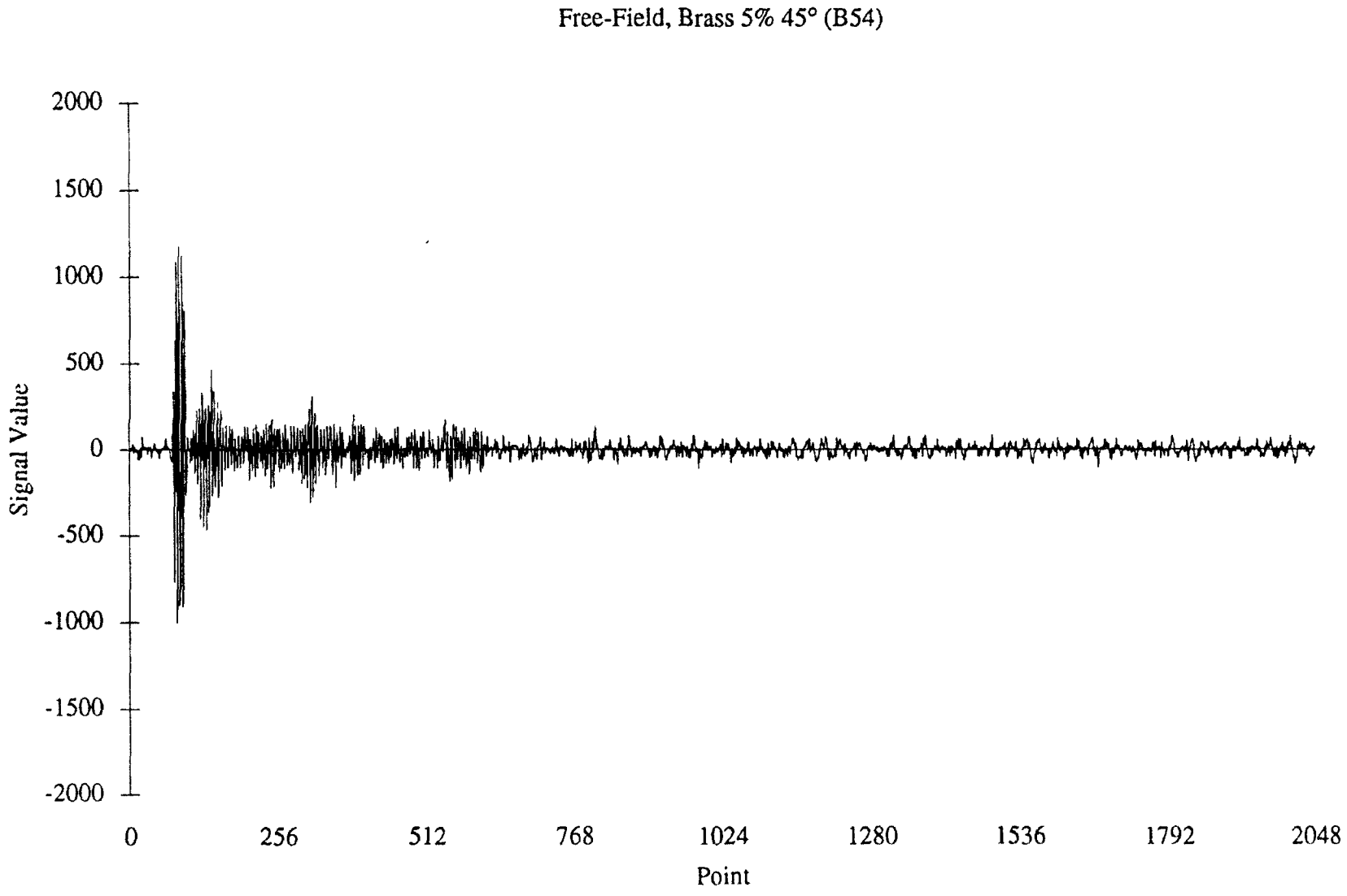
This is the first instance of each class of the free-field signals in time domain. The signals are in original form, but have had any DC offset removed.

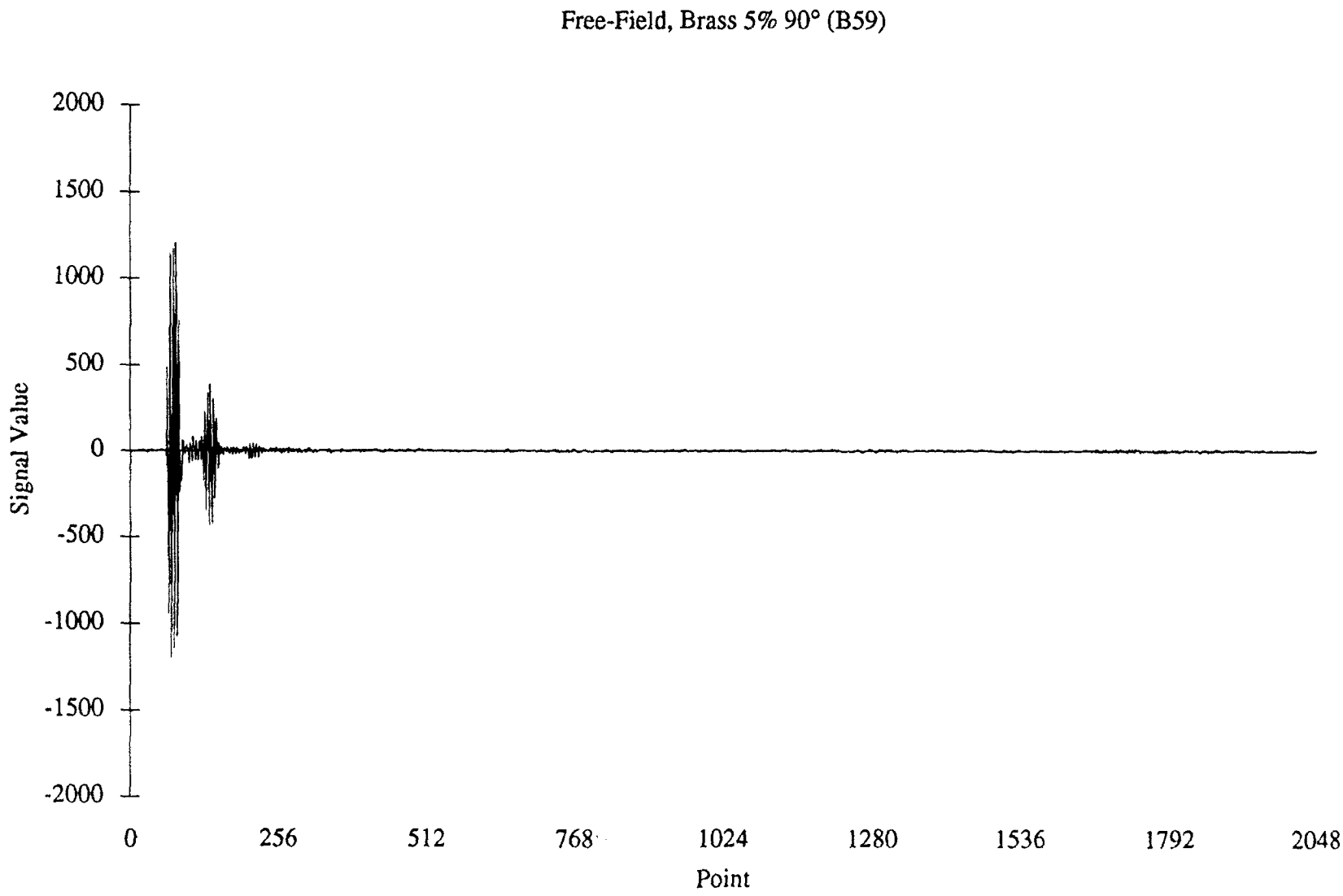

Free-Field, Brass 10% 0° (B10)
Signal Value
2000
1500
1000
500
0
-500
-1000
-1500
-2000
0
256
512
768
1024
1280
1536
1792
2048
Point

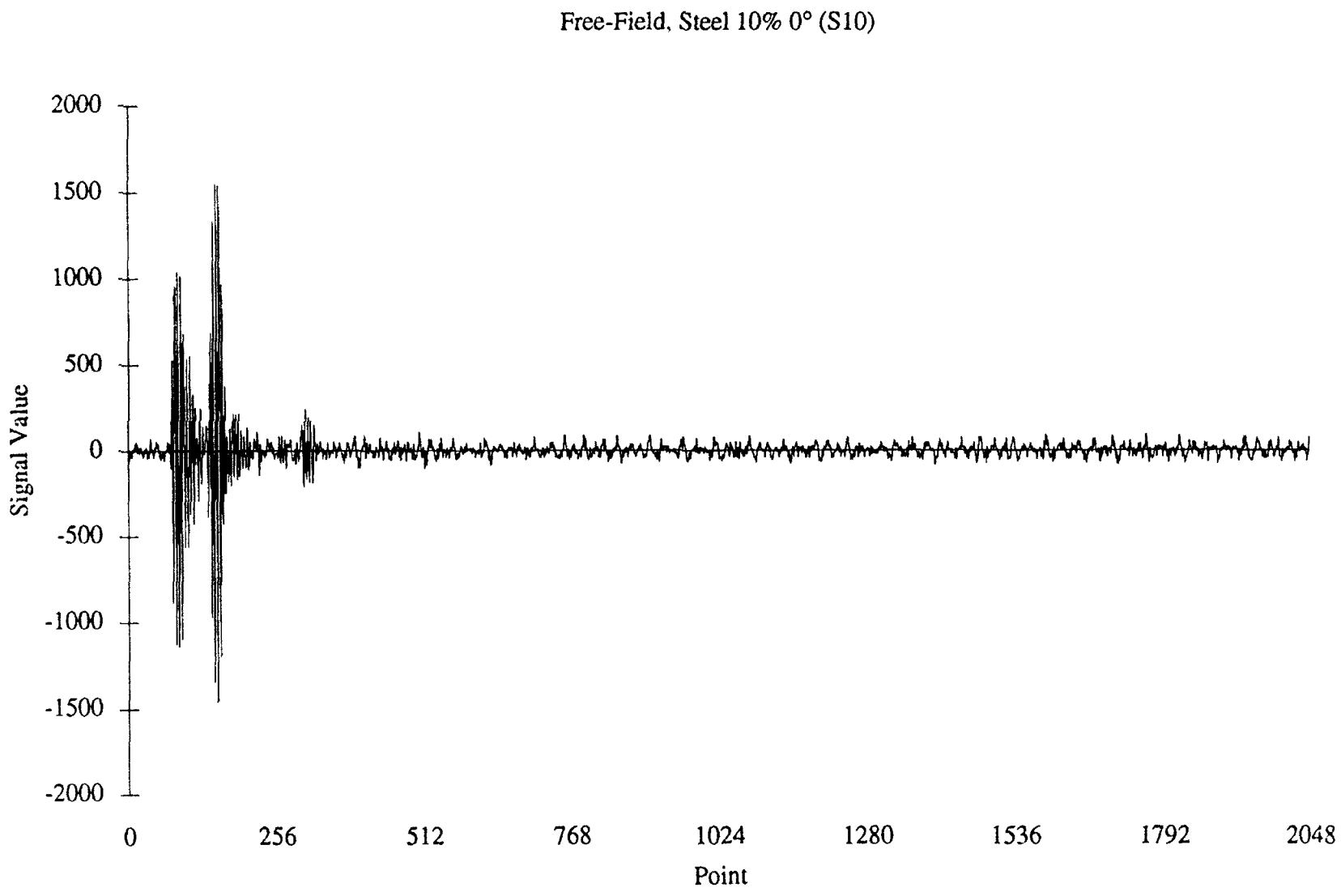

Free-Field, Brass 10% 45° (B14)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

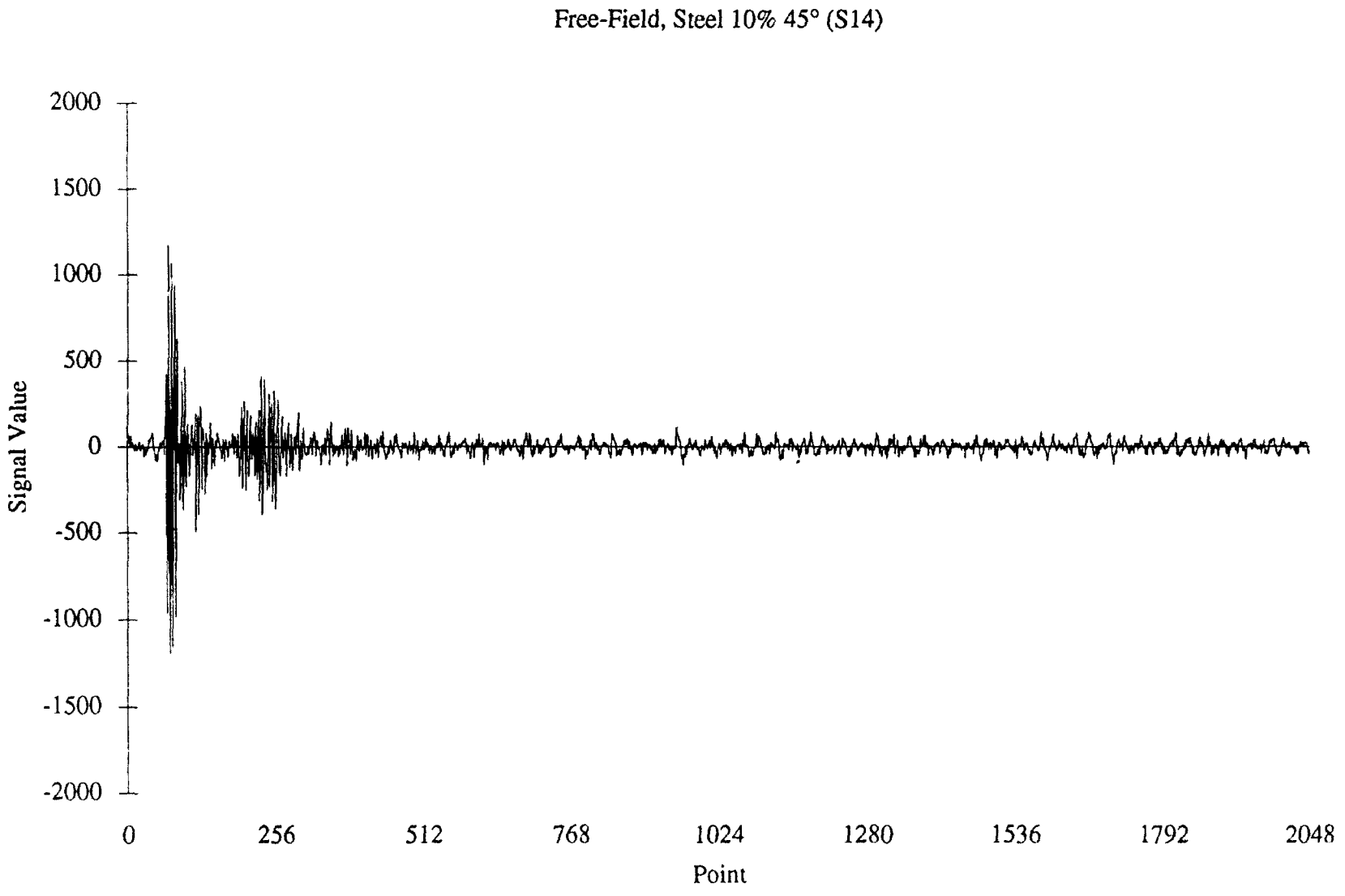

Free-Field, Brass 10% 90° (B19)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

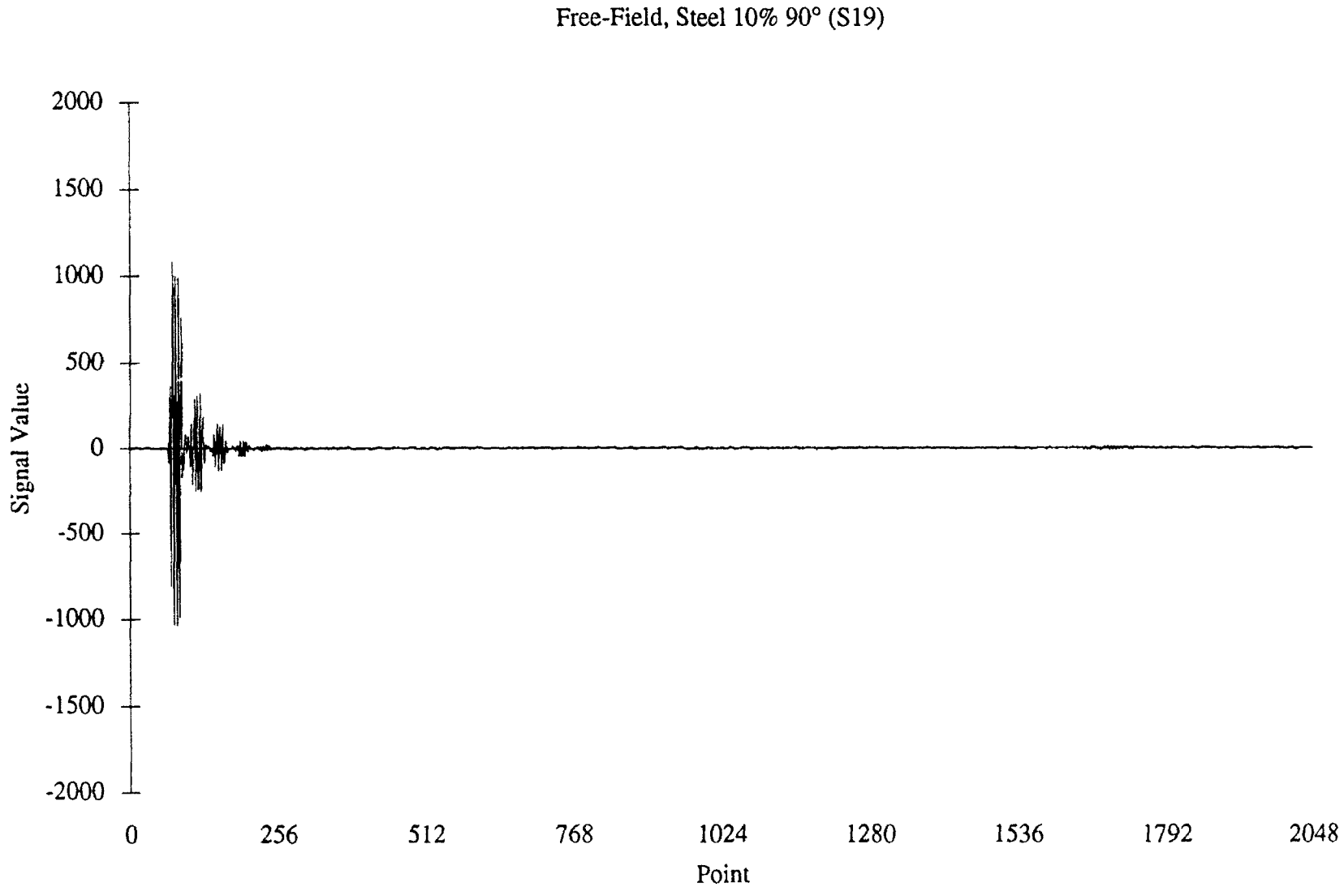

Free-Field, Brass 5% 0° (B50)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

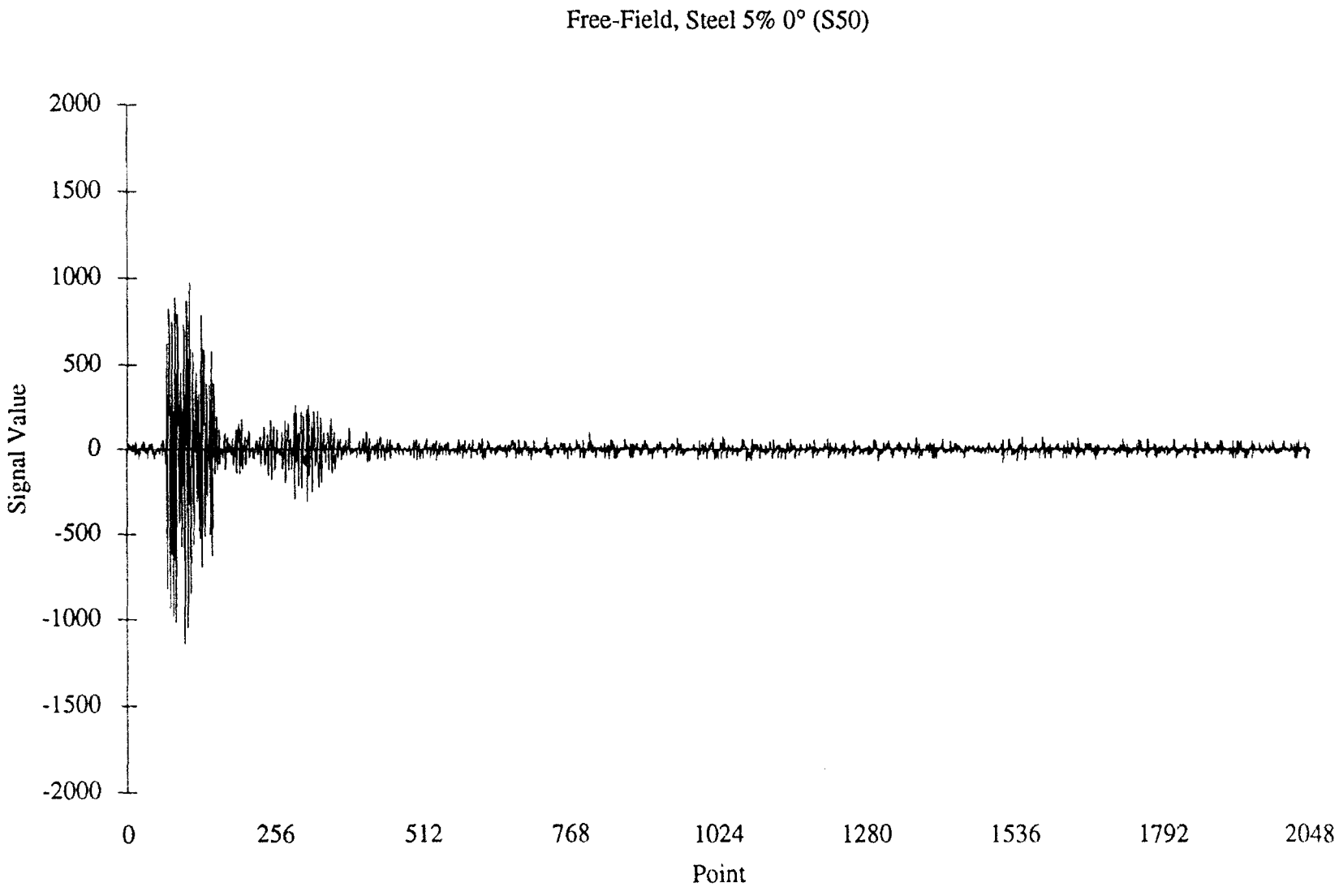

Free-Field, Brass 5% 45° (B54)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

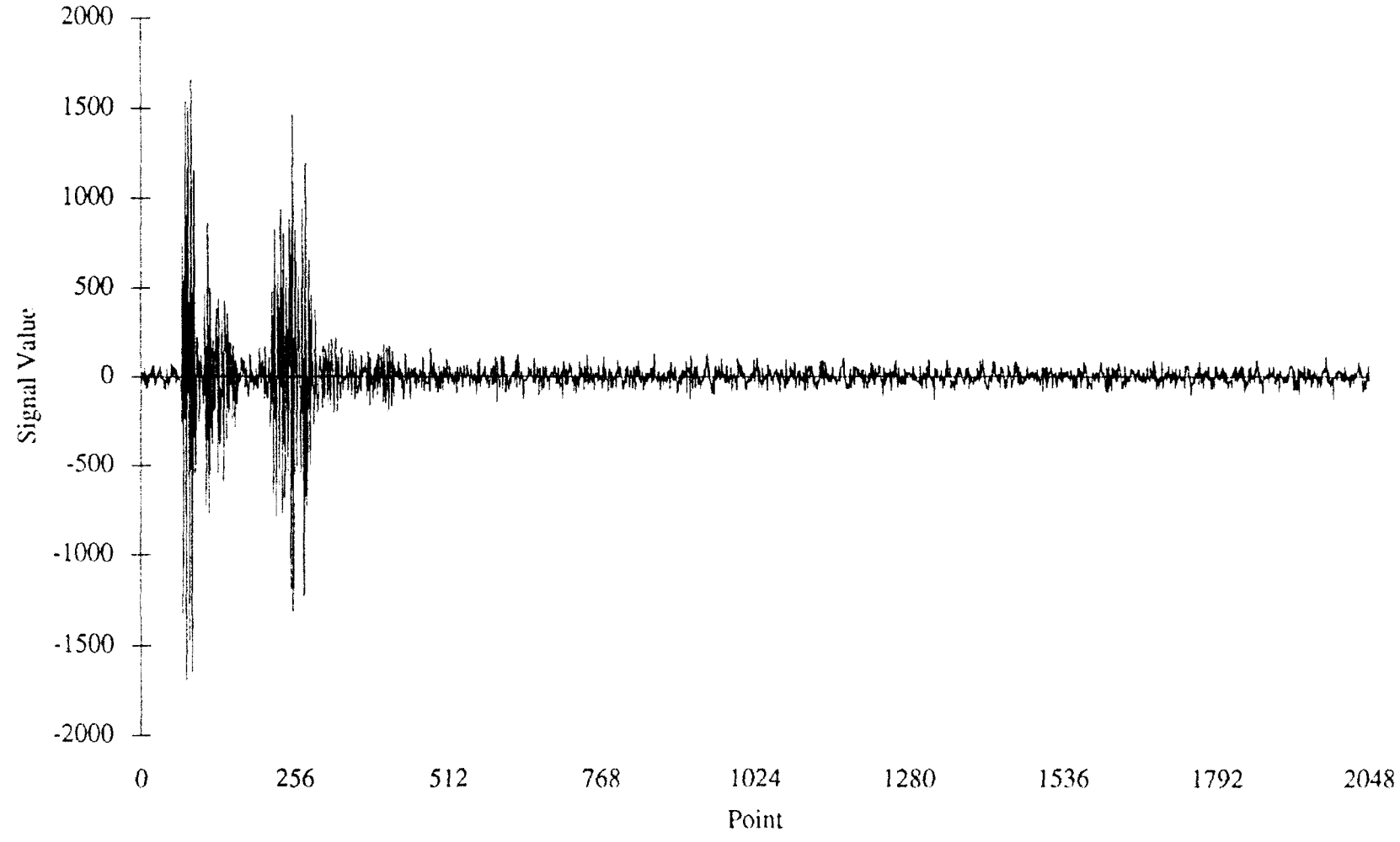

Free-Field, Brass 5% 90° (B59)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

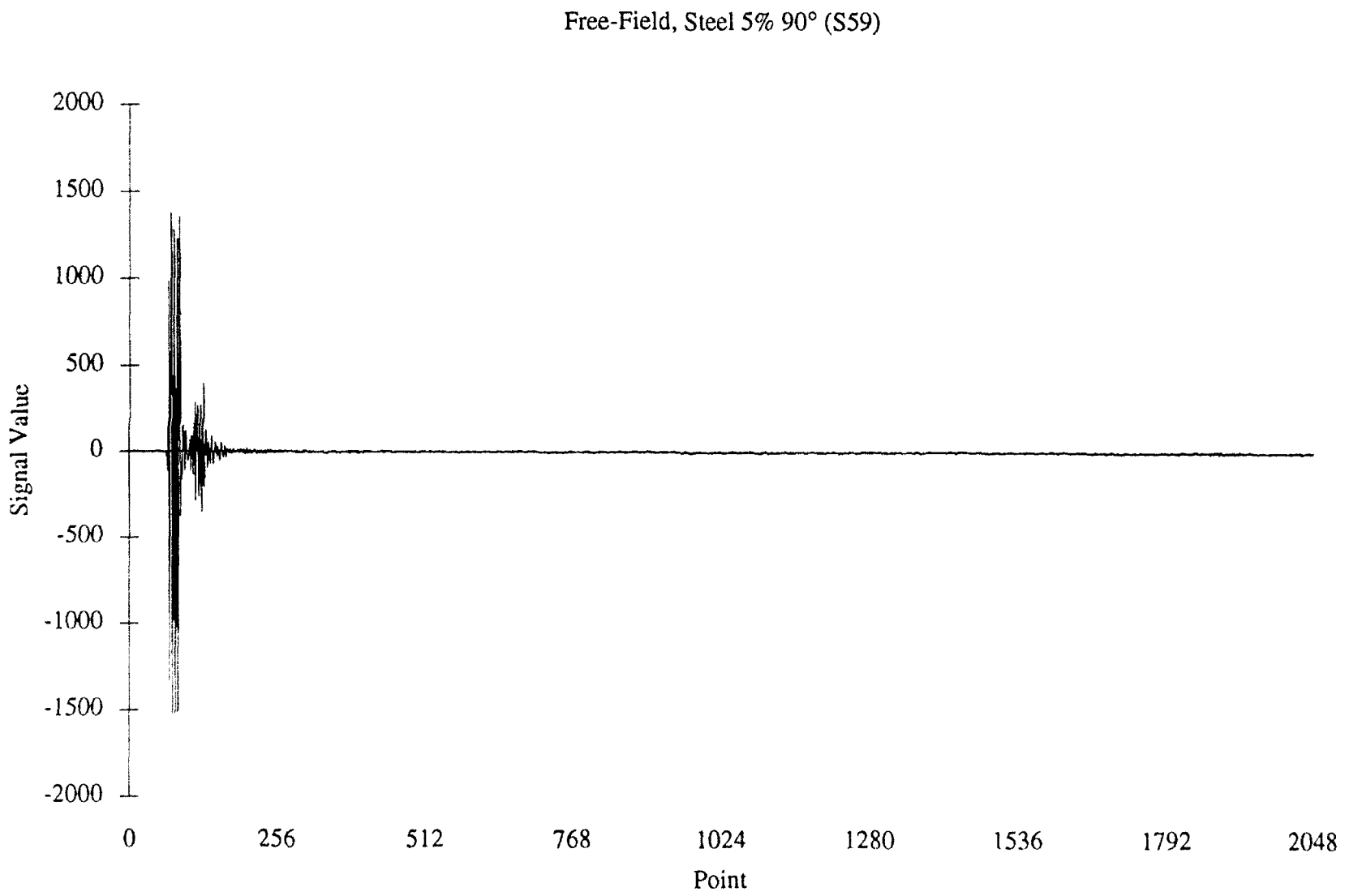

Free-Field, Steel 10% 0° (S10)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point


Free-Field, Steel 10% 45° (S14)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point


Free-Field, Steel 10% 90° (S19)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point


Free-Field, Steel 5% 0° (S50)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

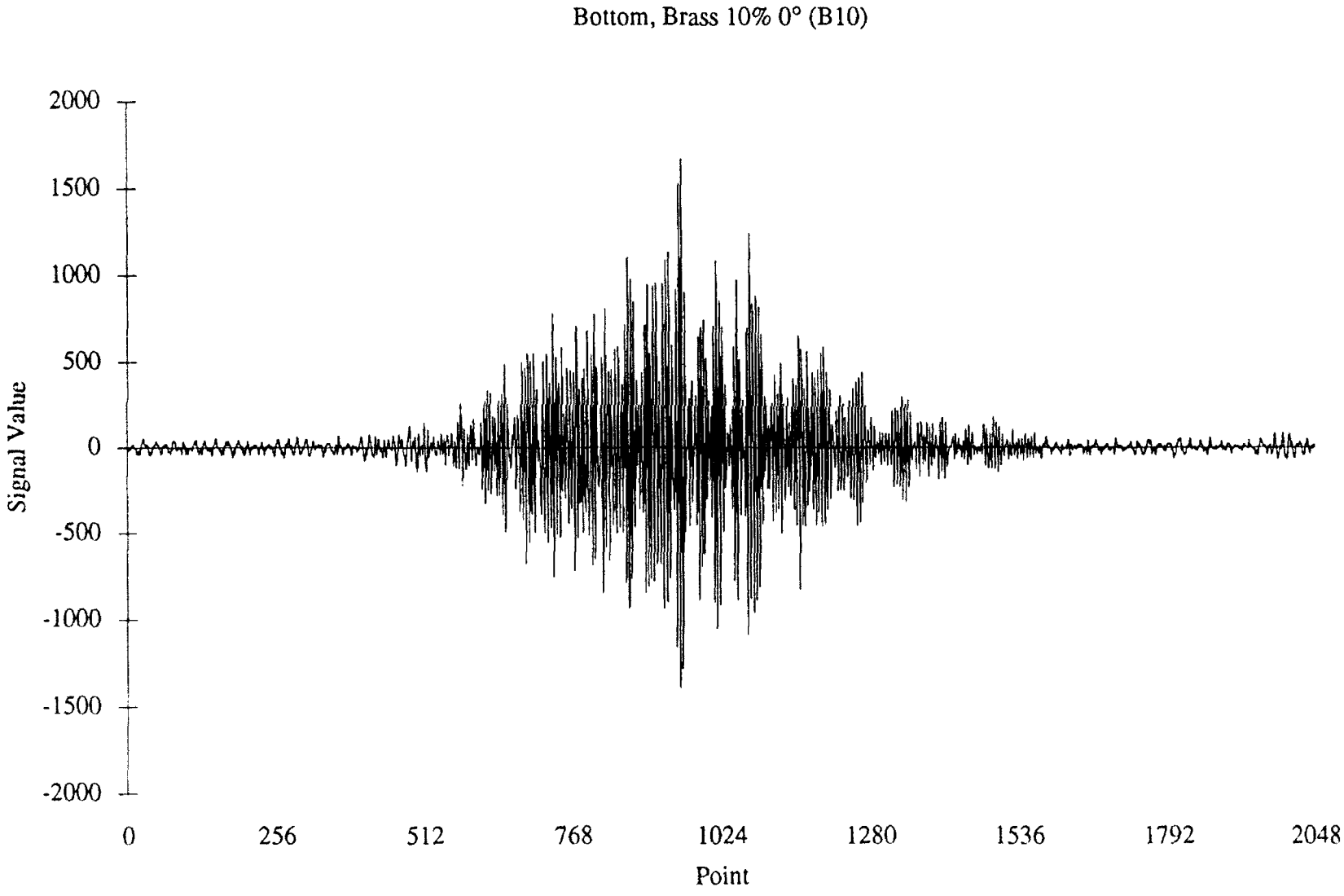

Free-Field, Steel 5% 45° (S54)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

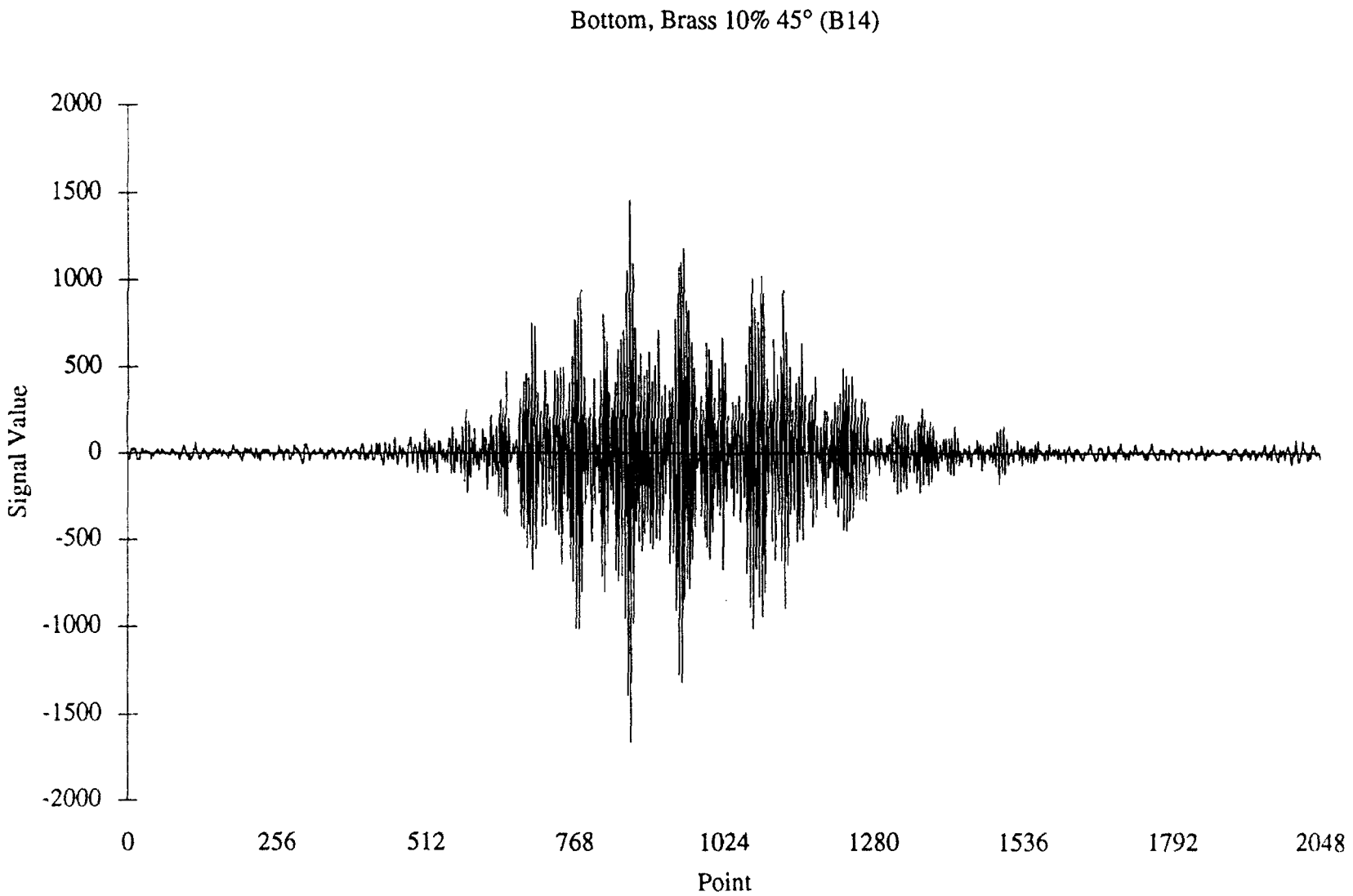

Free-Field, Steel 5% 90° (S59)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

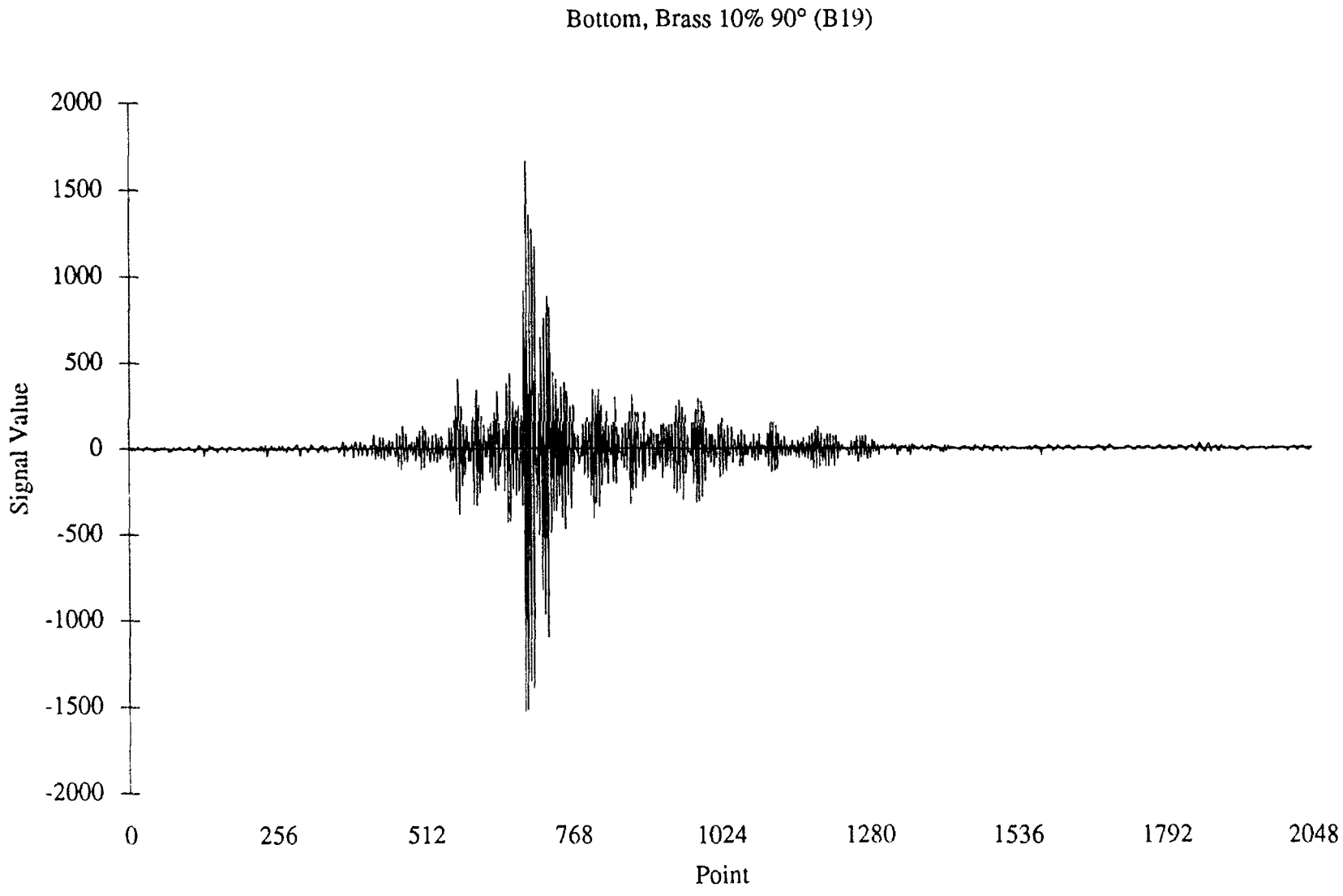
APPENDIX B

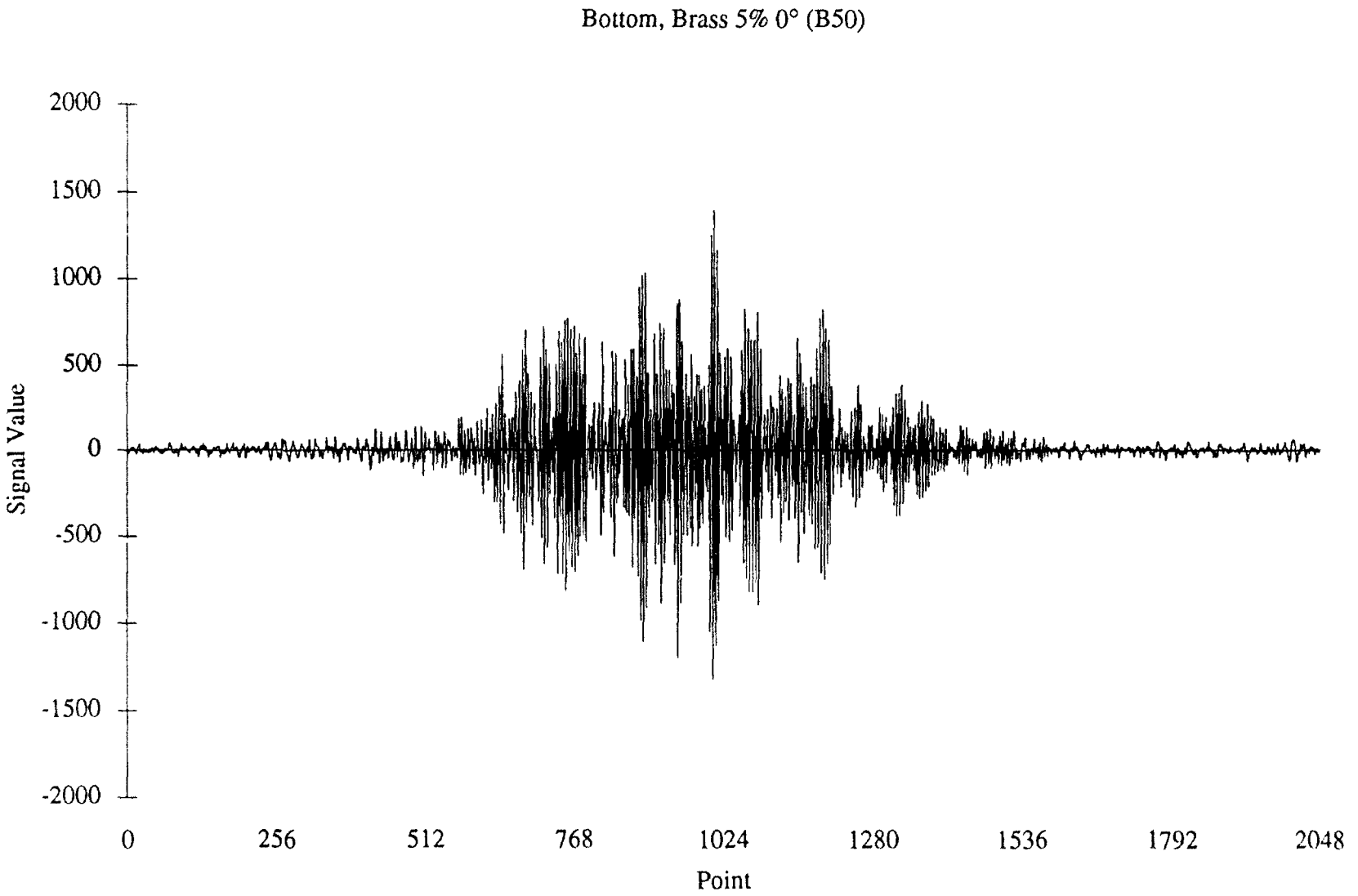
BOTTOM SIGNALS

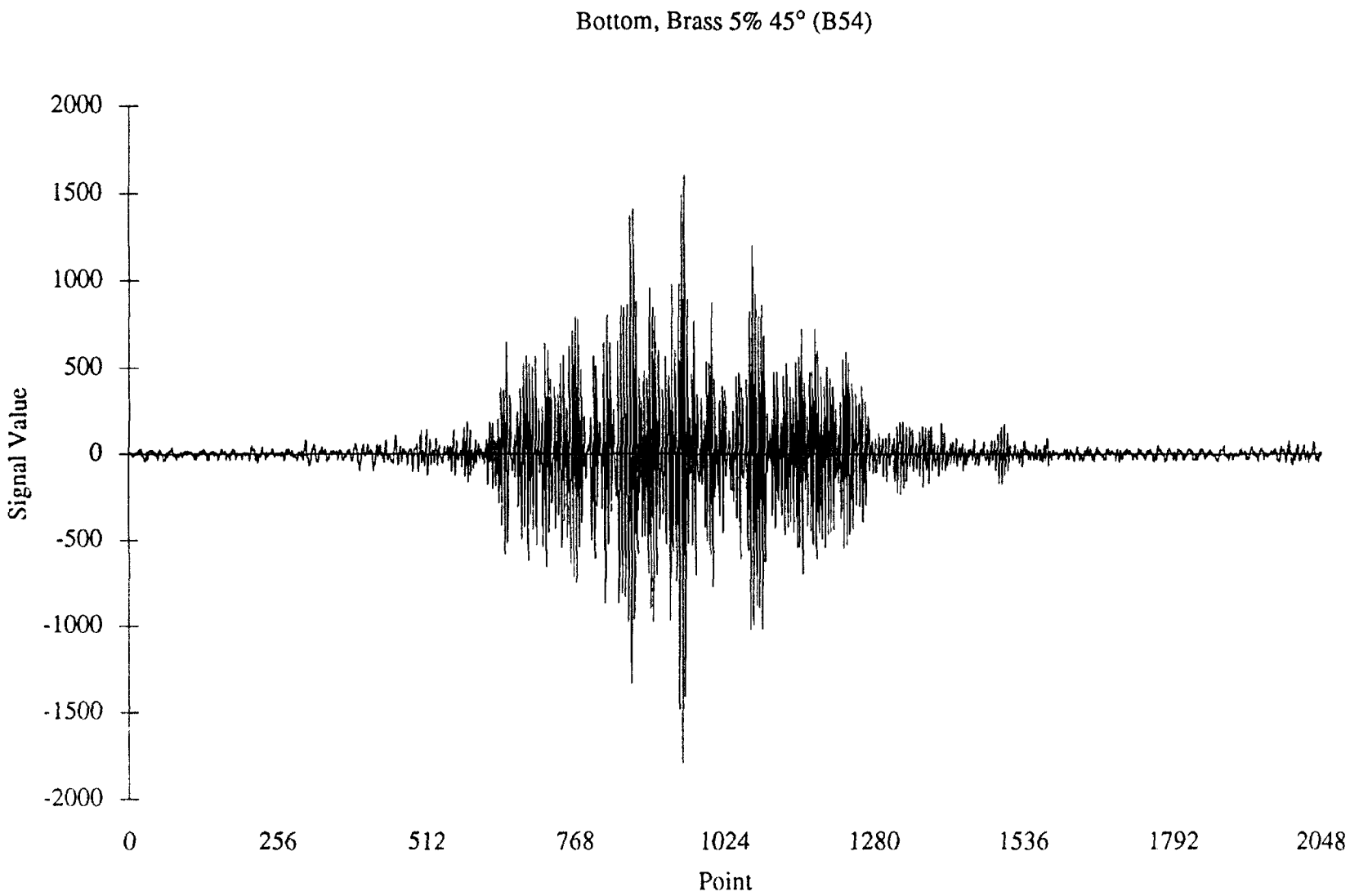
This is the first instance of each class of the bottom signals in time domain. The signals are in original form, but have had any DC offset removed.

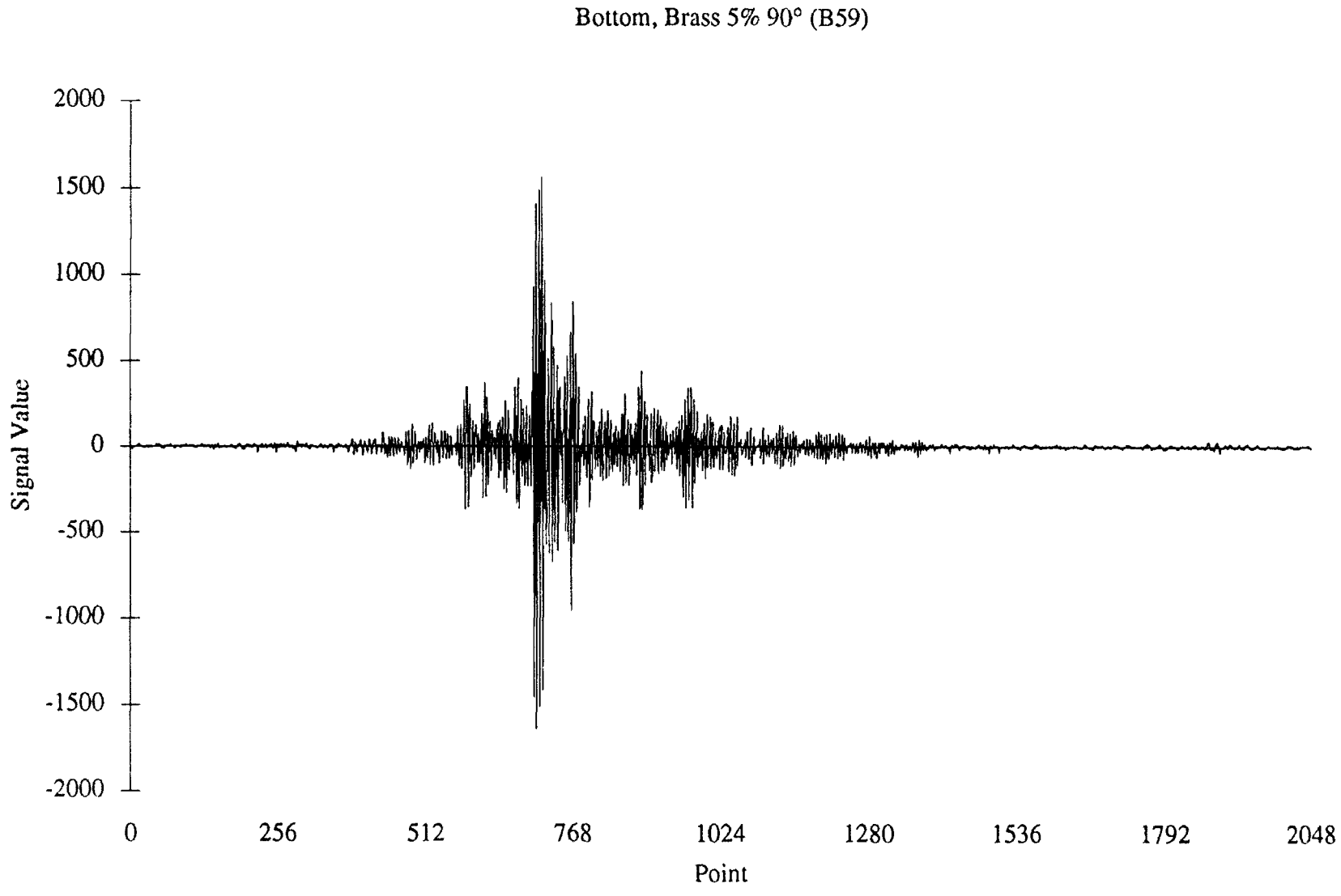

Bottom, Brass 10% 0° (B10)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

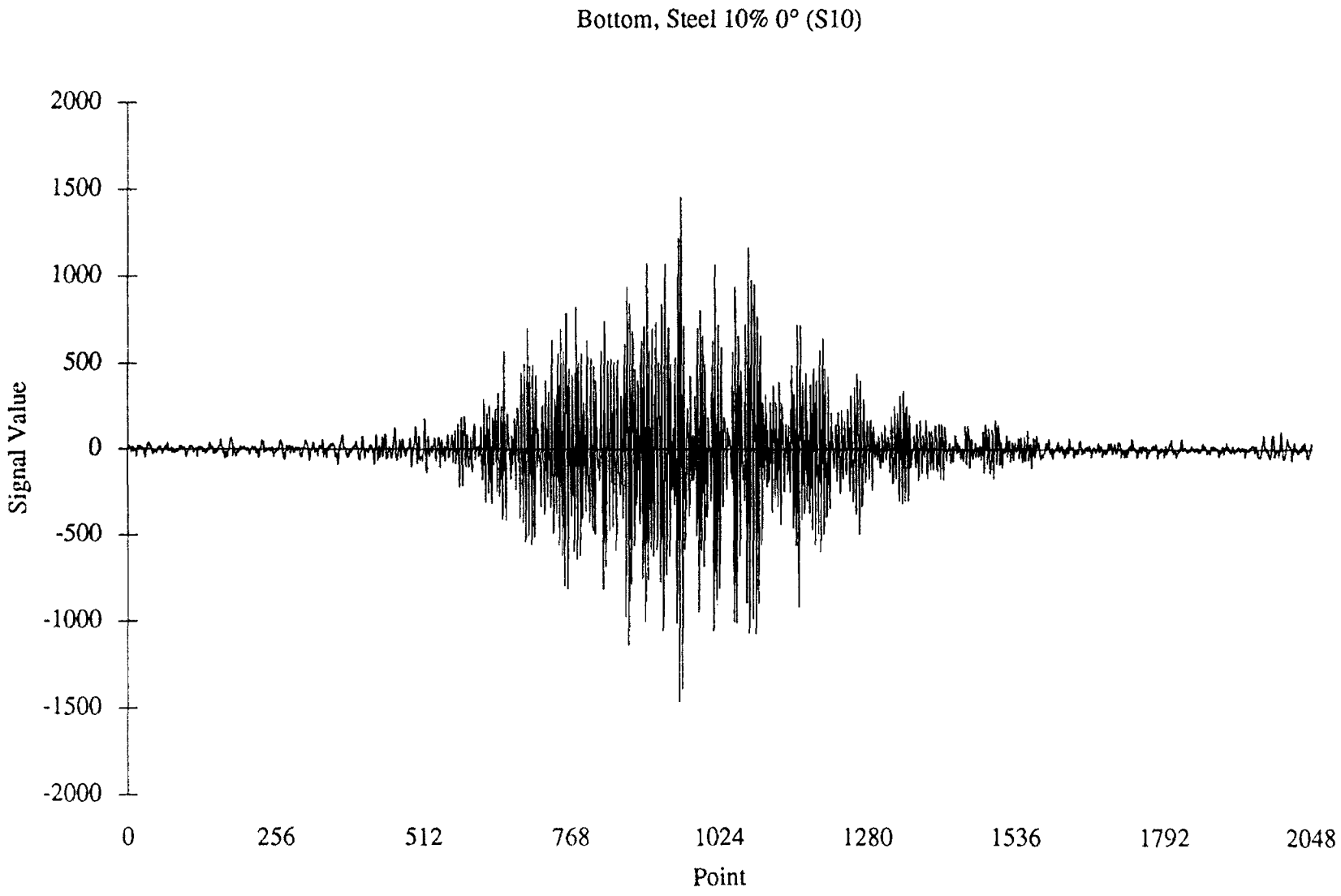

Bottom, Brass 10% 45° (B14)
Signal Value
2000
1500
1000
500
0
-500
-1000
-1500
-2000
0
256
512
768
1024
1280
1536
1792
2048
Point

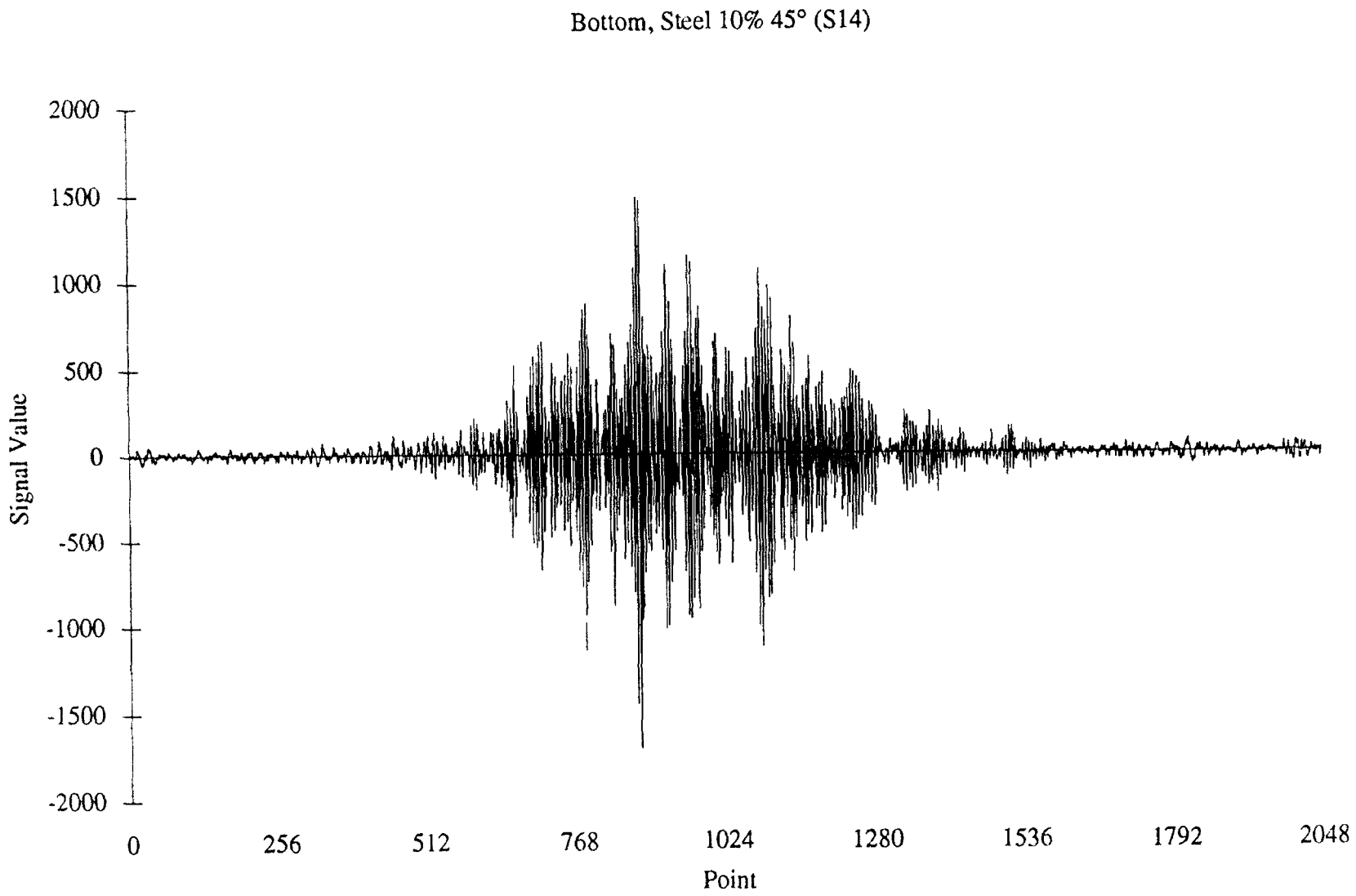

Bottom, Brass 10% 90° (B19)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

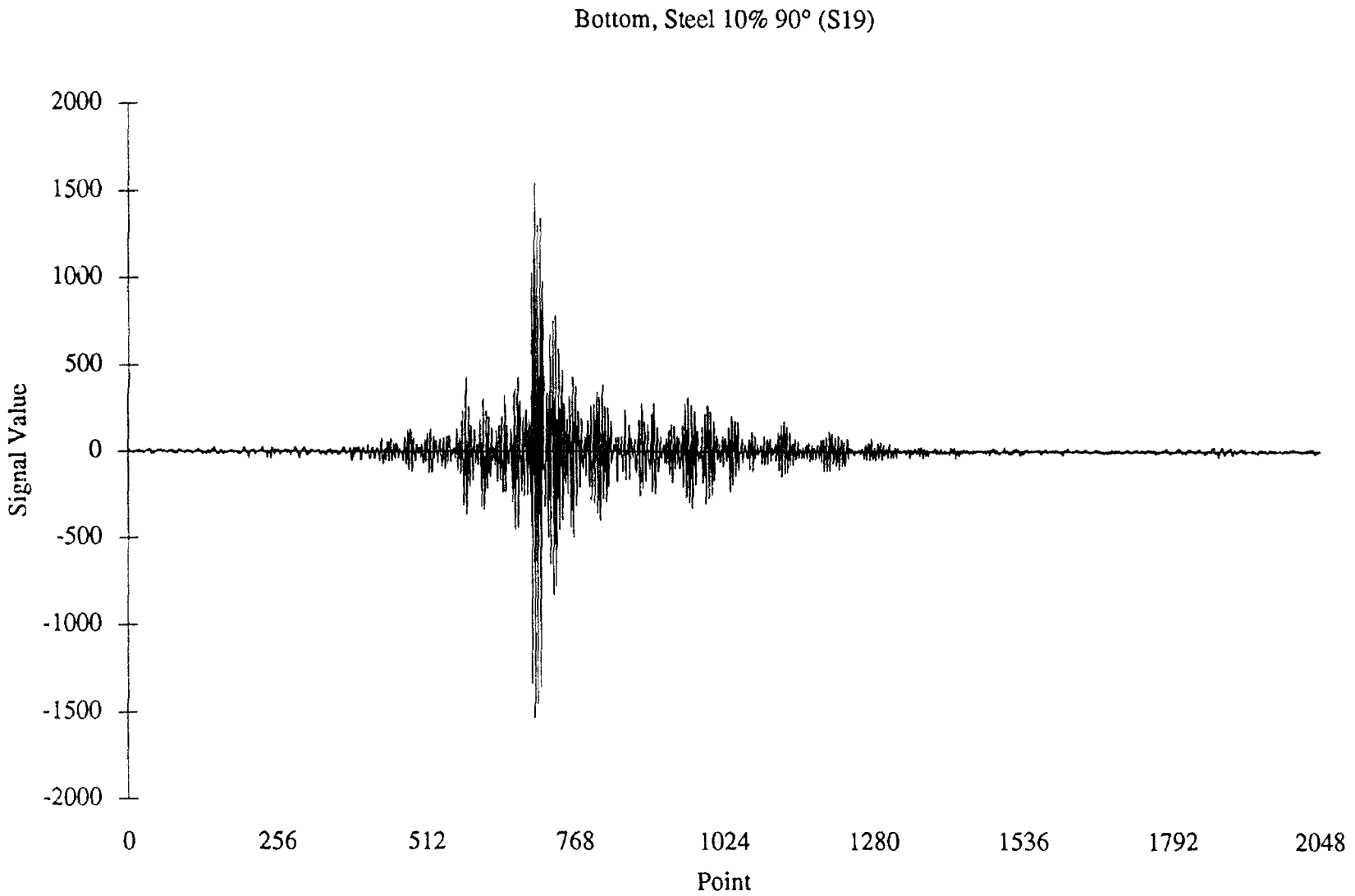

Bottom, Brass 5% 0° (B50)
Signal Value
2000
1500
1000
500
0
-500
-1000
-1500
-2000
0
256
512
768
1024
1280
1536
1792
2048
Point

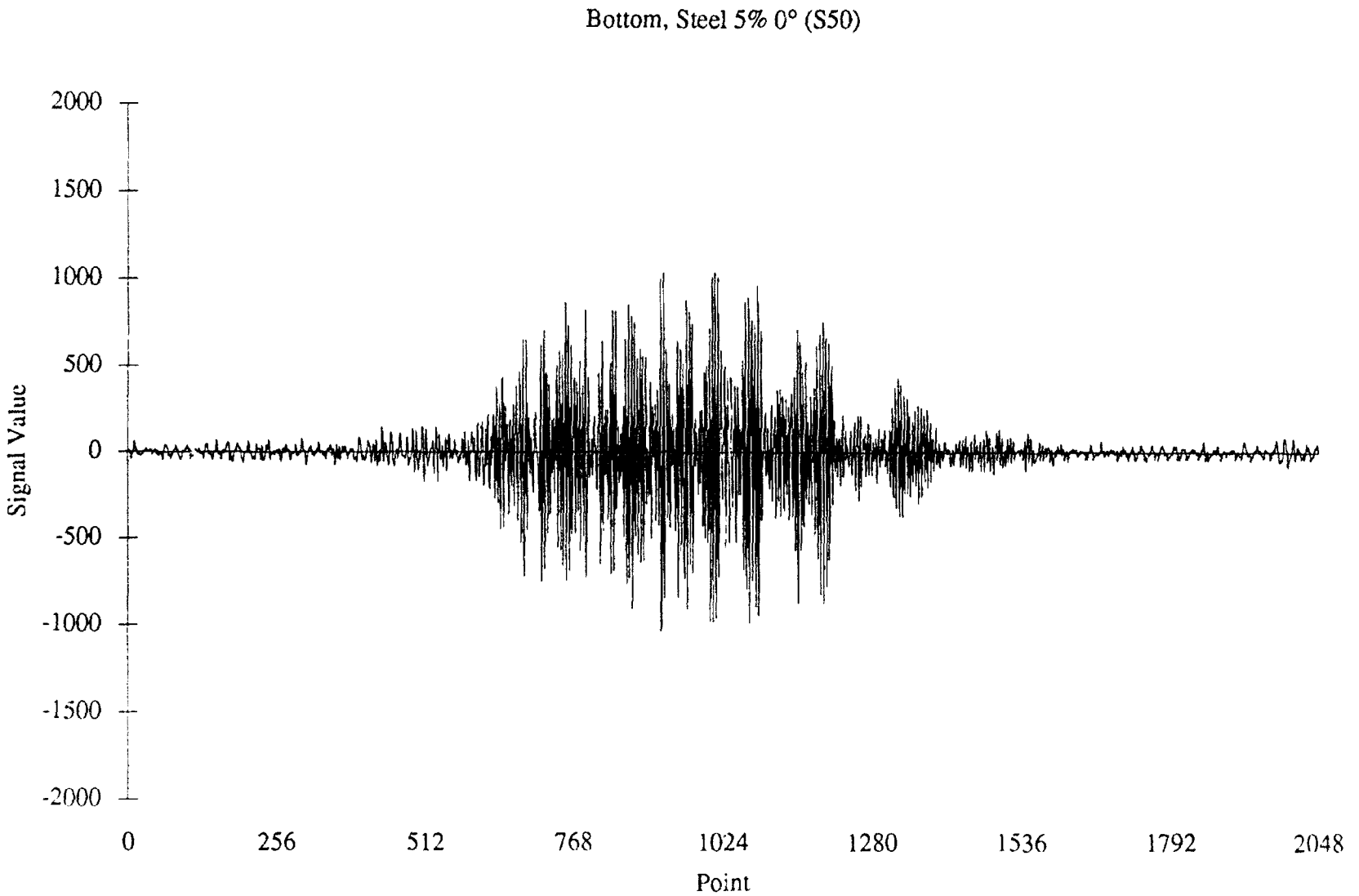

Bottom, Brass 5% 45° (B54)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

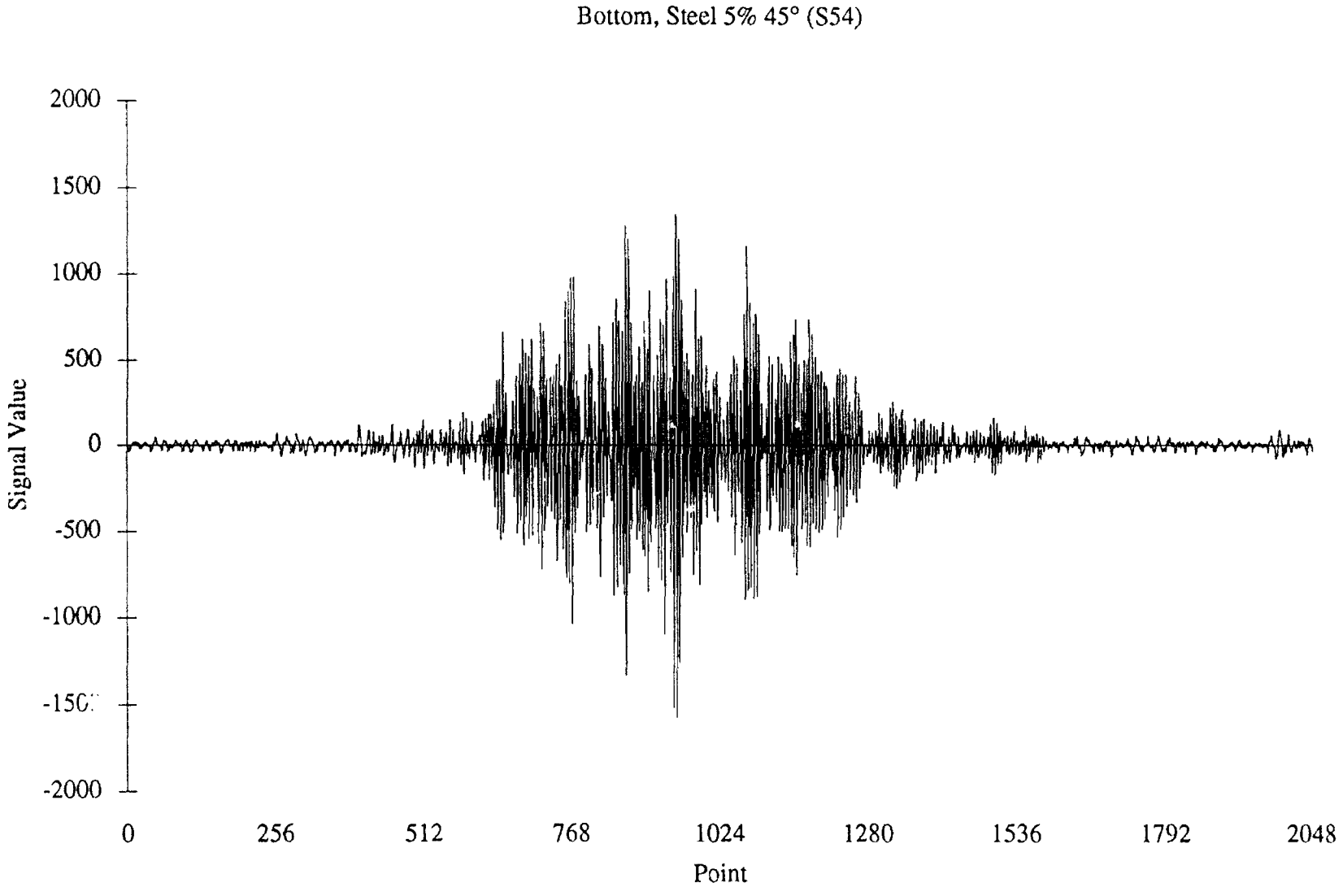

Bottom, Brass 5% 90° (B59)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

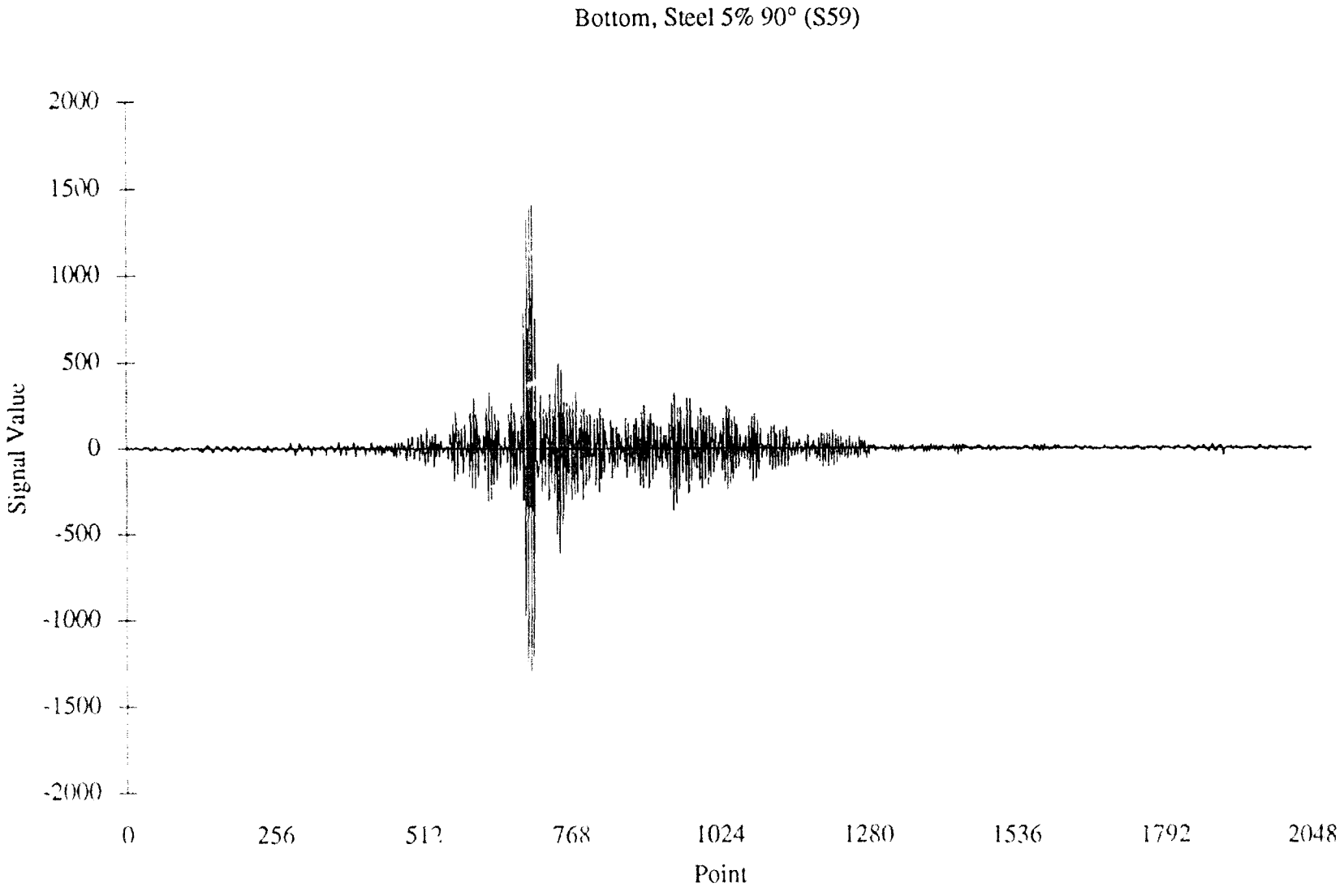

Bottom, Steel 10% 0° (S10)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point


Bottom, Steel 10% 45° (S14)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point


Bottom, Steel 10% 90° (S19)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point


Bottom, Steel 5% 0° (S50)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

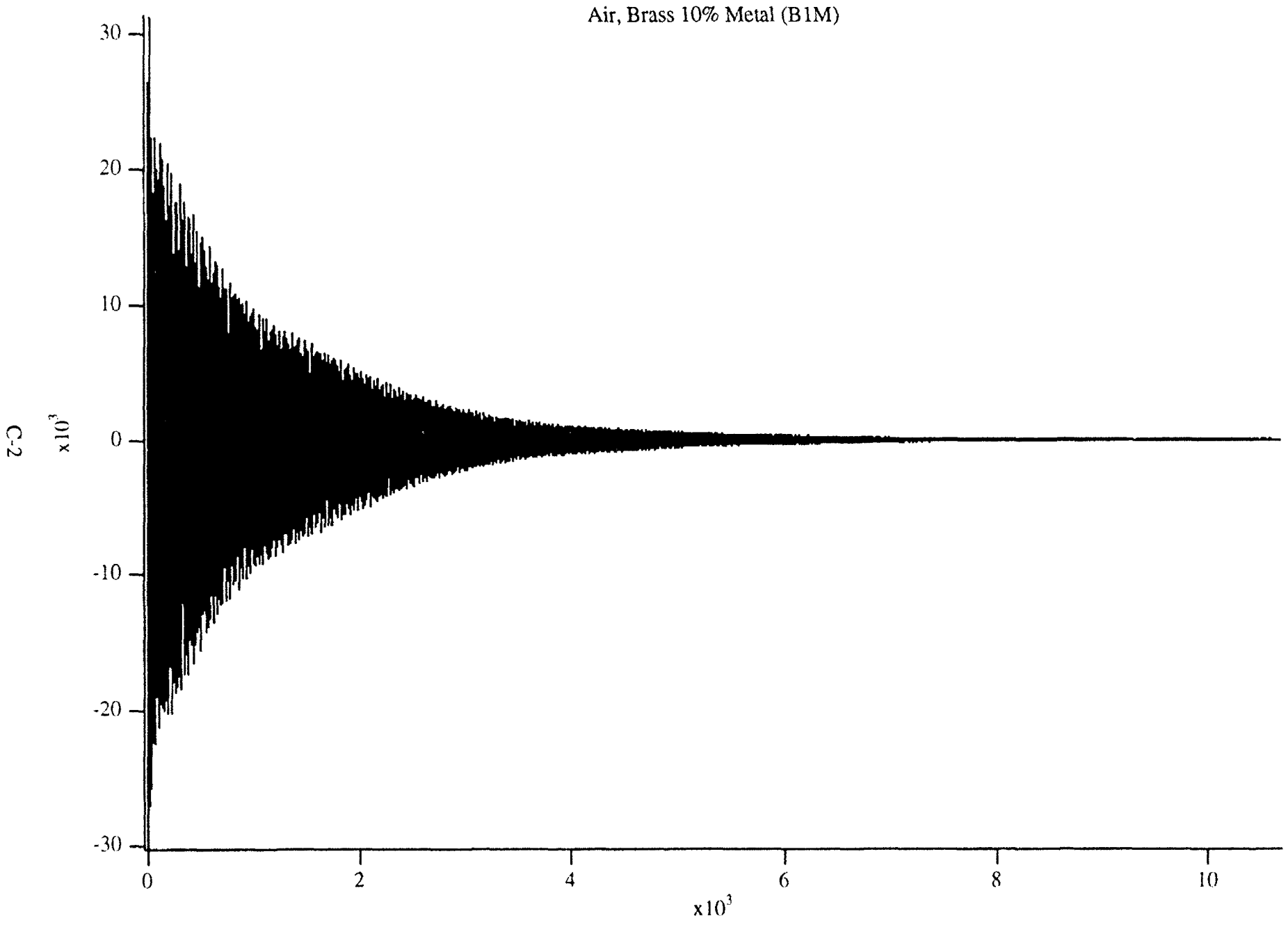

Bottom, Steel 5% 45° (S54)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

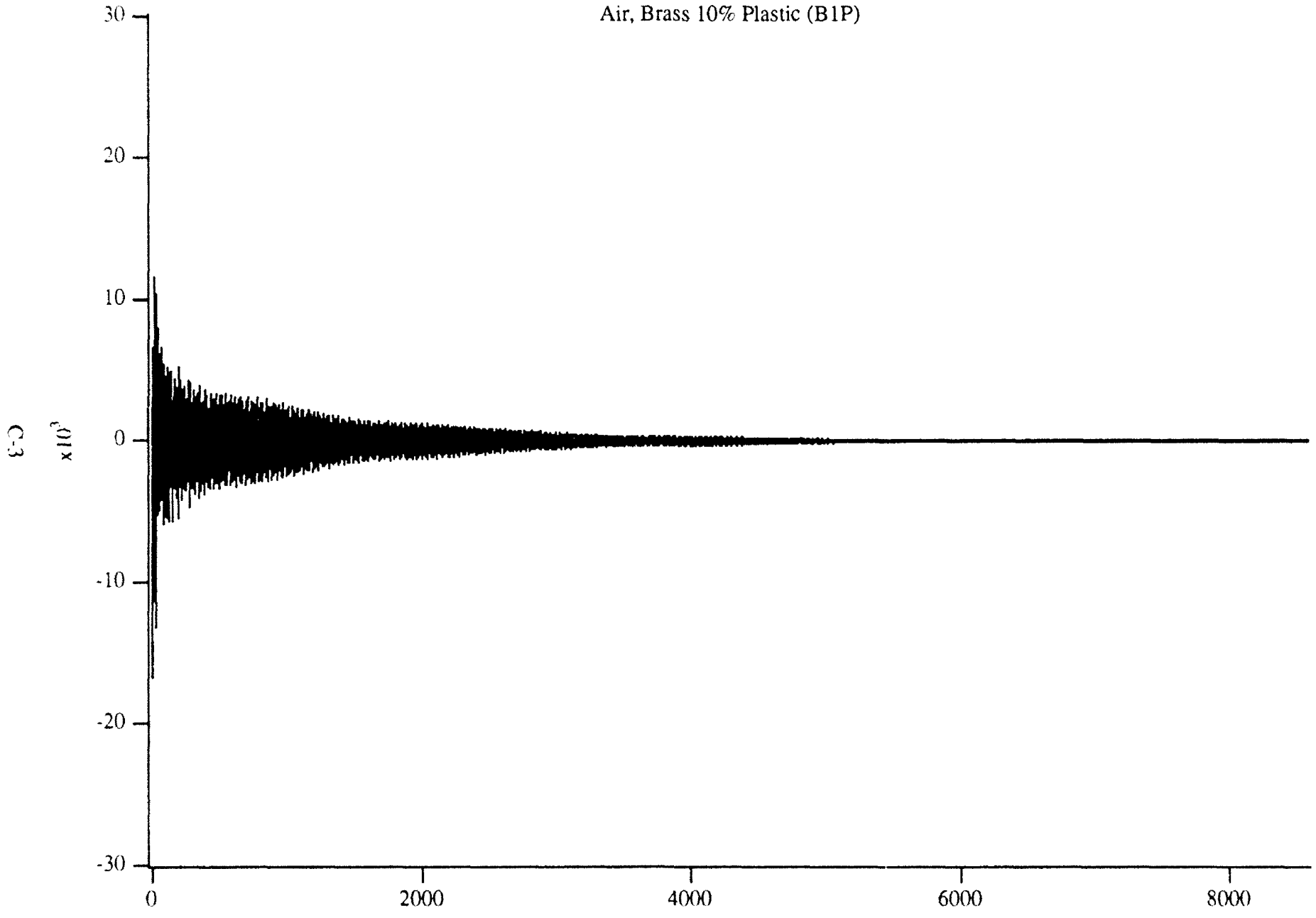

Bottom, Steel 5% 90° (S59)
2000
1500
1000
500
0
-500
-1000
-1500
-2000
Signal Value
0
256
512
768
1024
1280
1536
1792
2048
Point

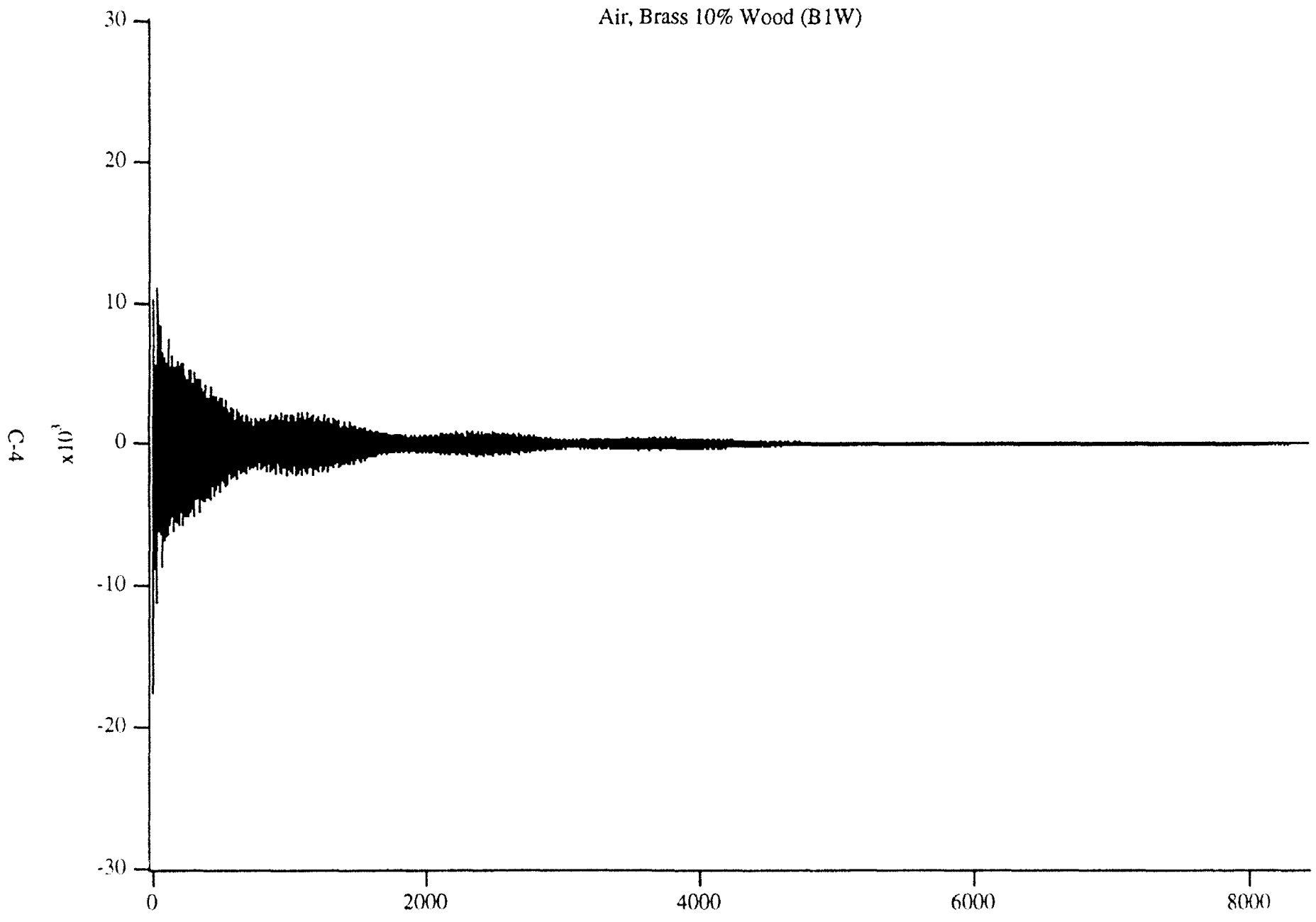
APPENDIX C

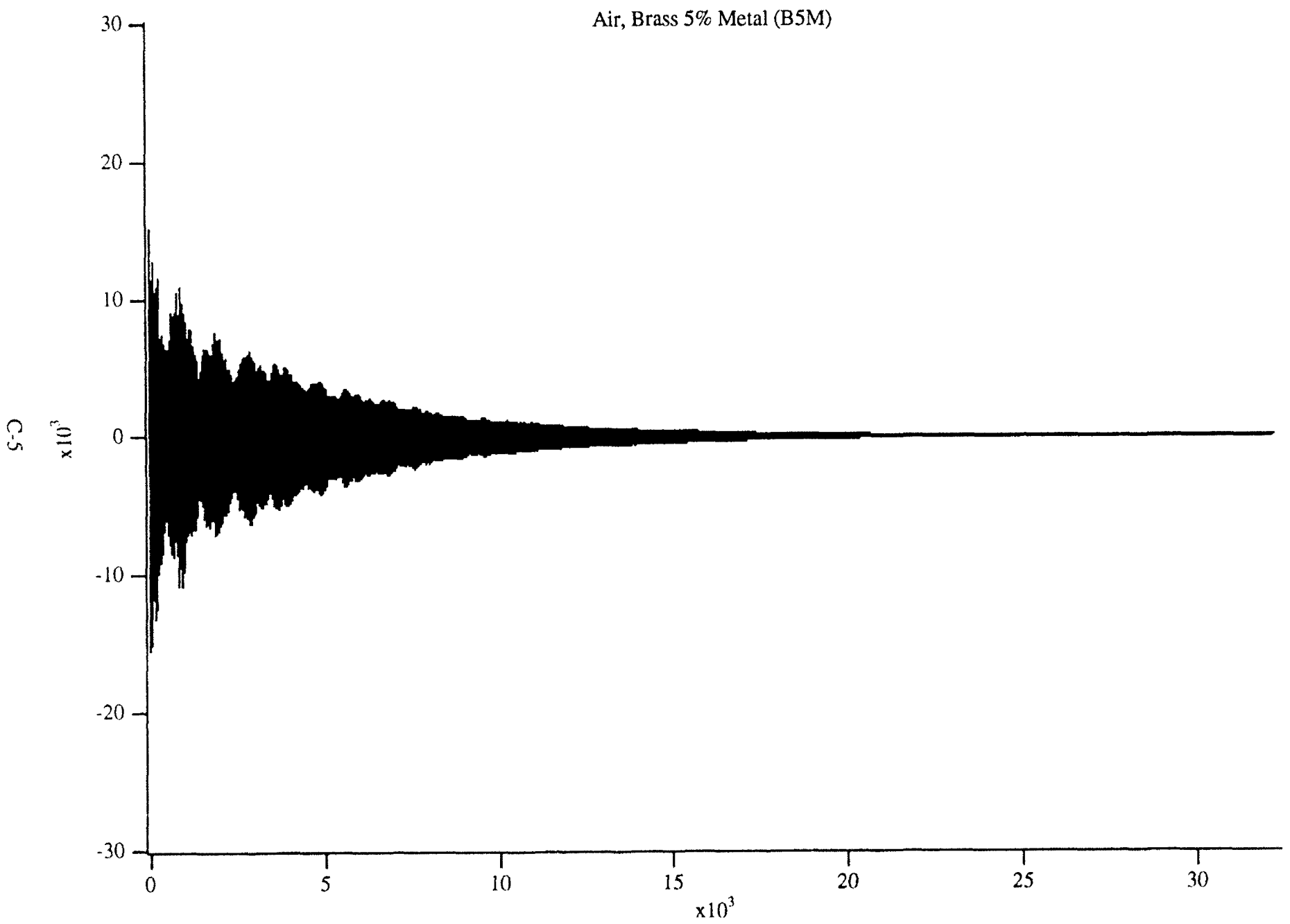
AIR SIGNALS

This is the first instance of each class of the air signals in time domain. The signals are in original form. Note that the scale for the x axis may differ between classes.

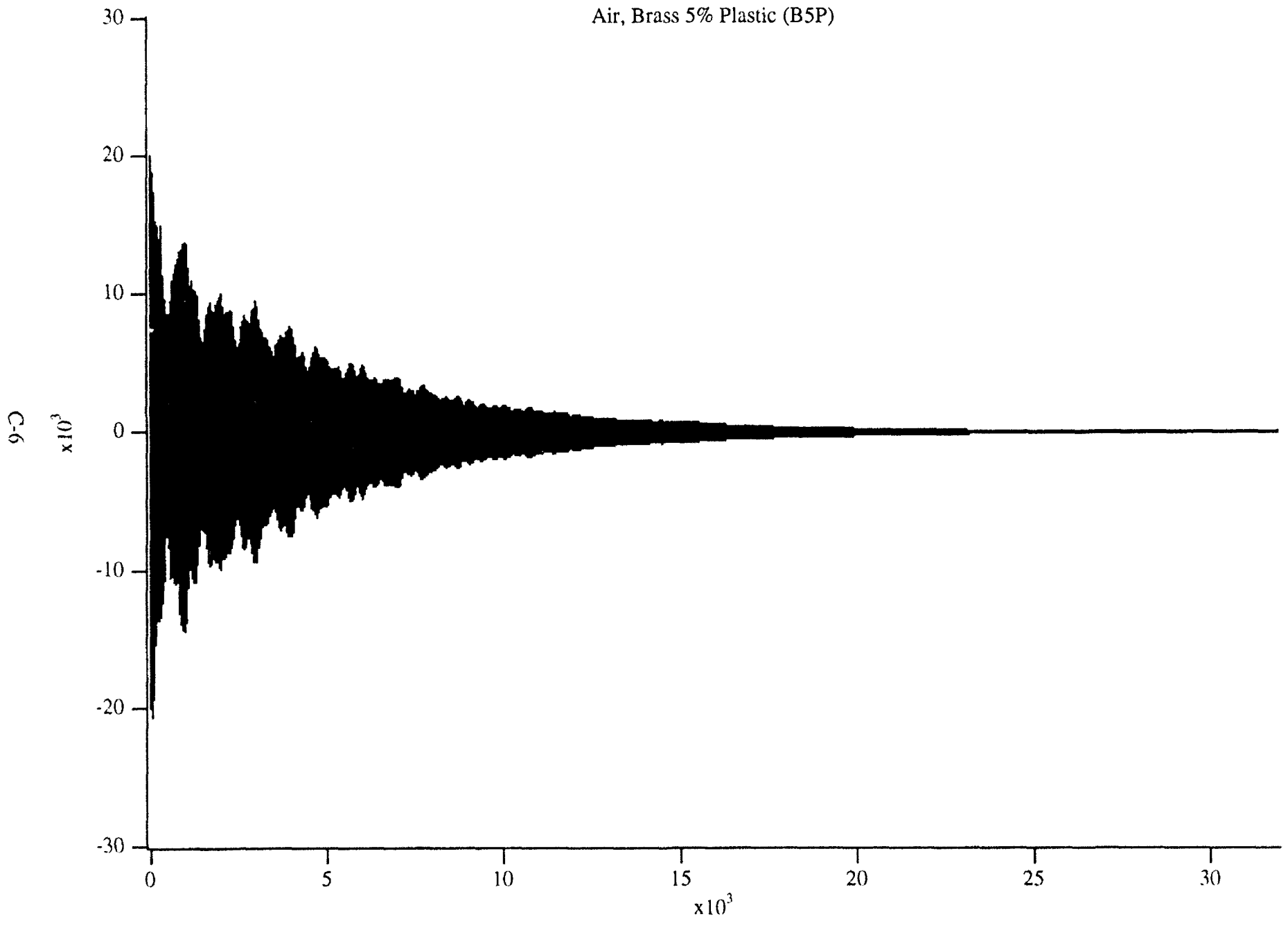

Air, Brass 10% Metal (B1M)
C-2
x10^3
30
20
10
0
-10
-20
-30
0
2
4
6
8
10
x10^3

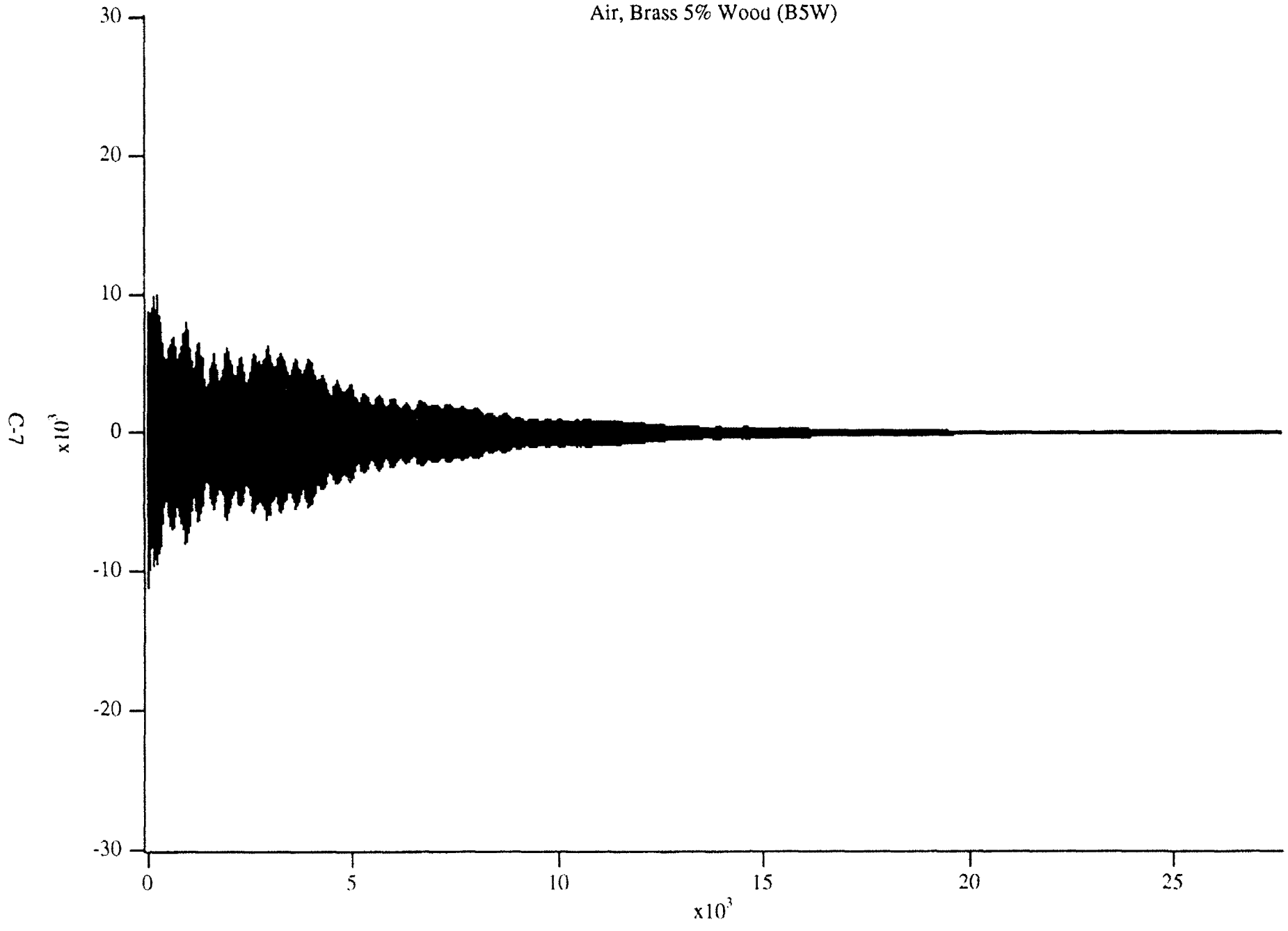

Air, Brass 10% Plastic (B1P)
C-3
x10^3
30
20
10
0
-10
-20
-30
0
2000
4000
6000
8000

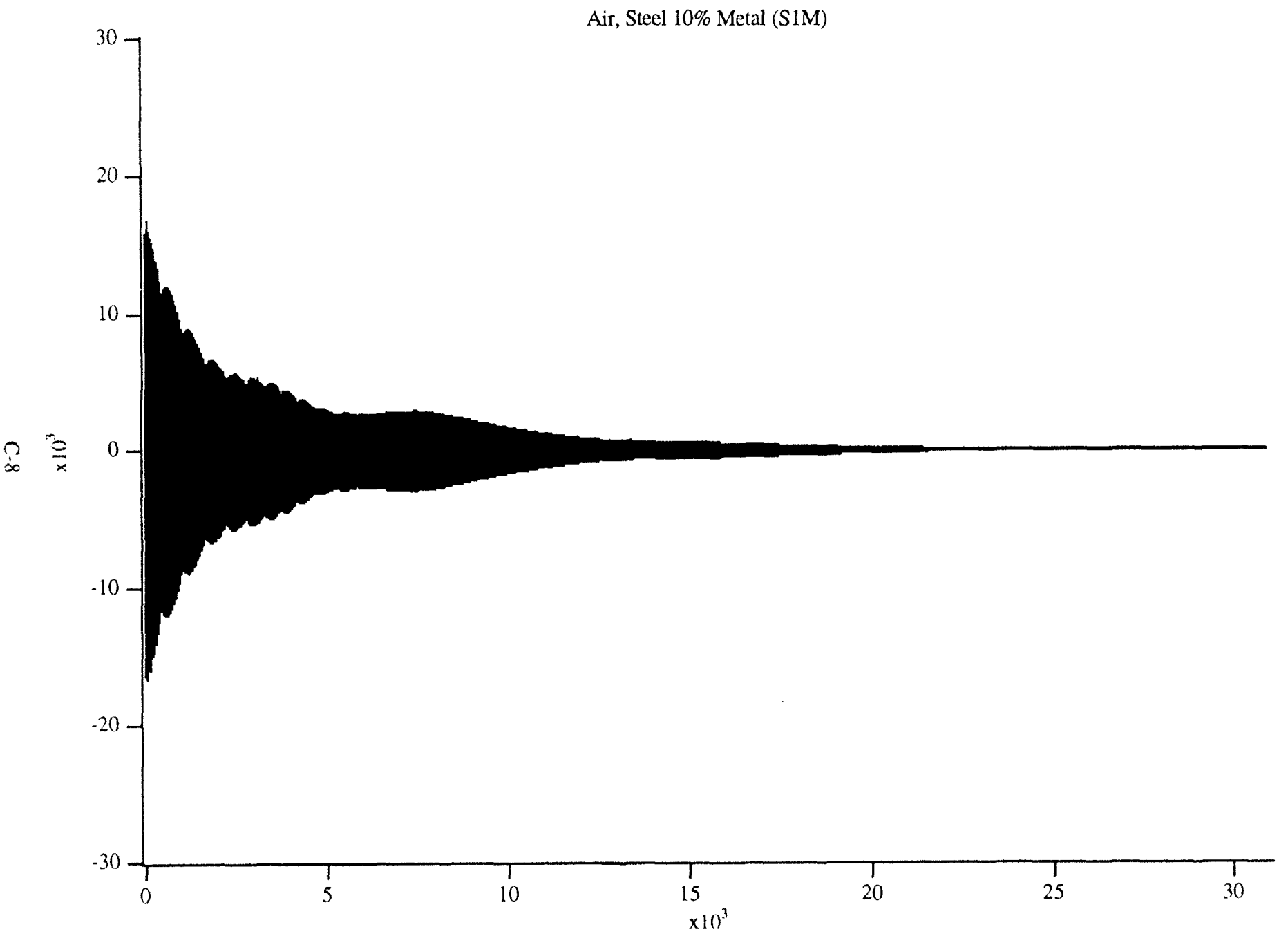

Air, Brass 10% Wood (B1W)
C-4
x10^3
30
20
10
0
-10
-20
-30
0
2000
4000
6000
8000

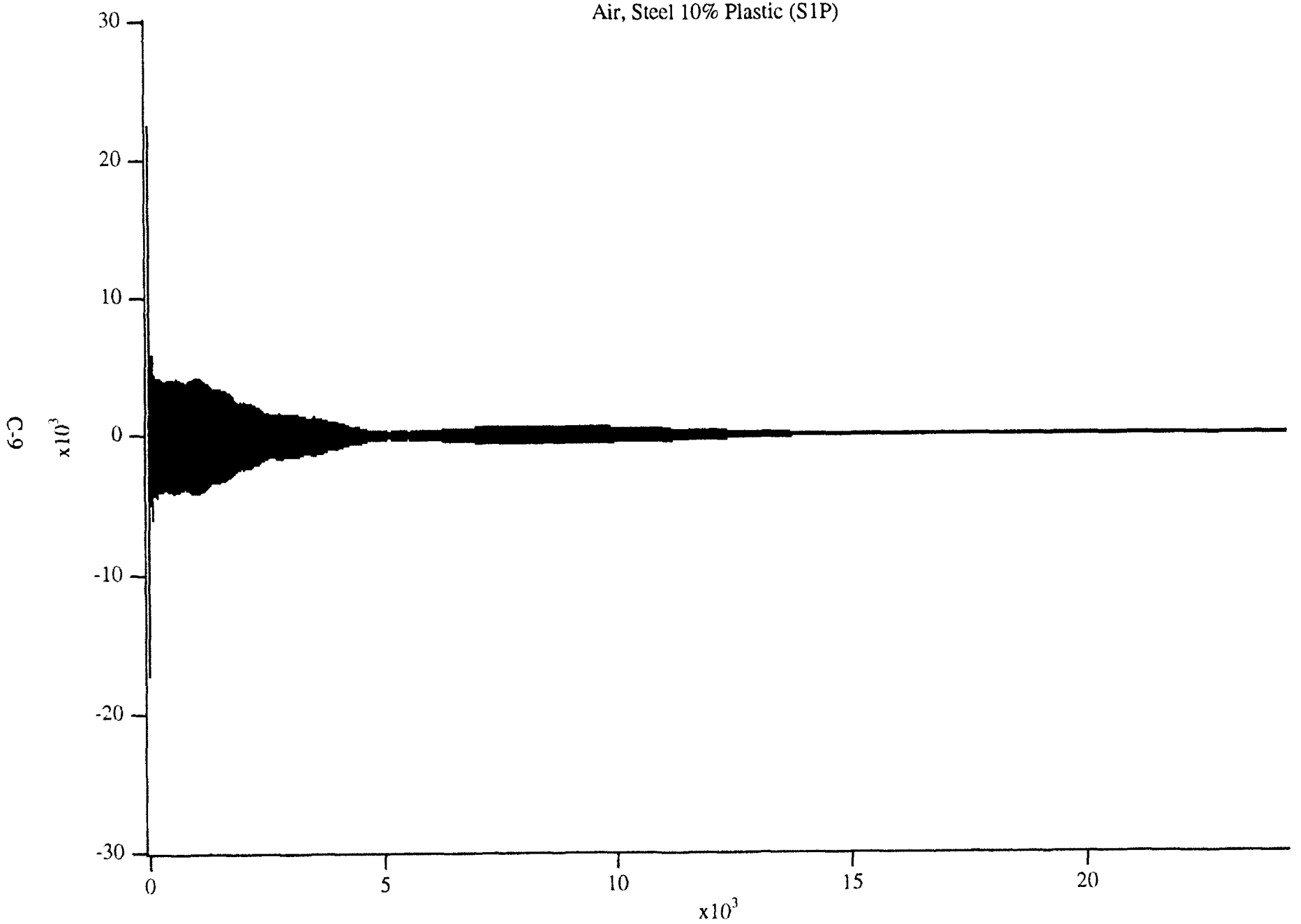

Air, Brass 5% Metal (B5M)
30
20
10
0
-10
-20
-30
x10^3
0
5
10
15
20
25
30
x10^3

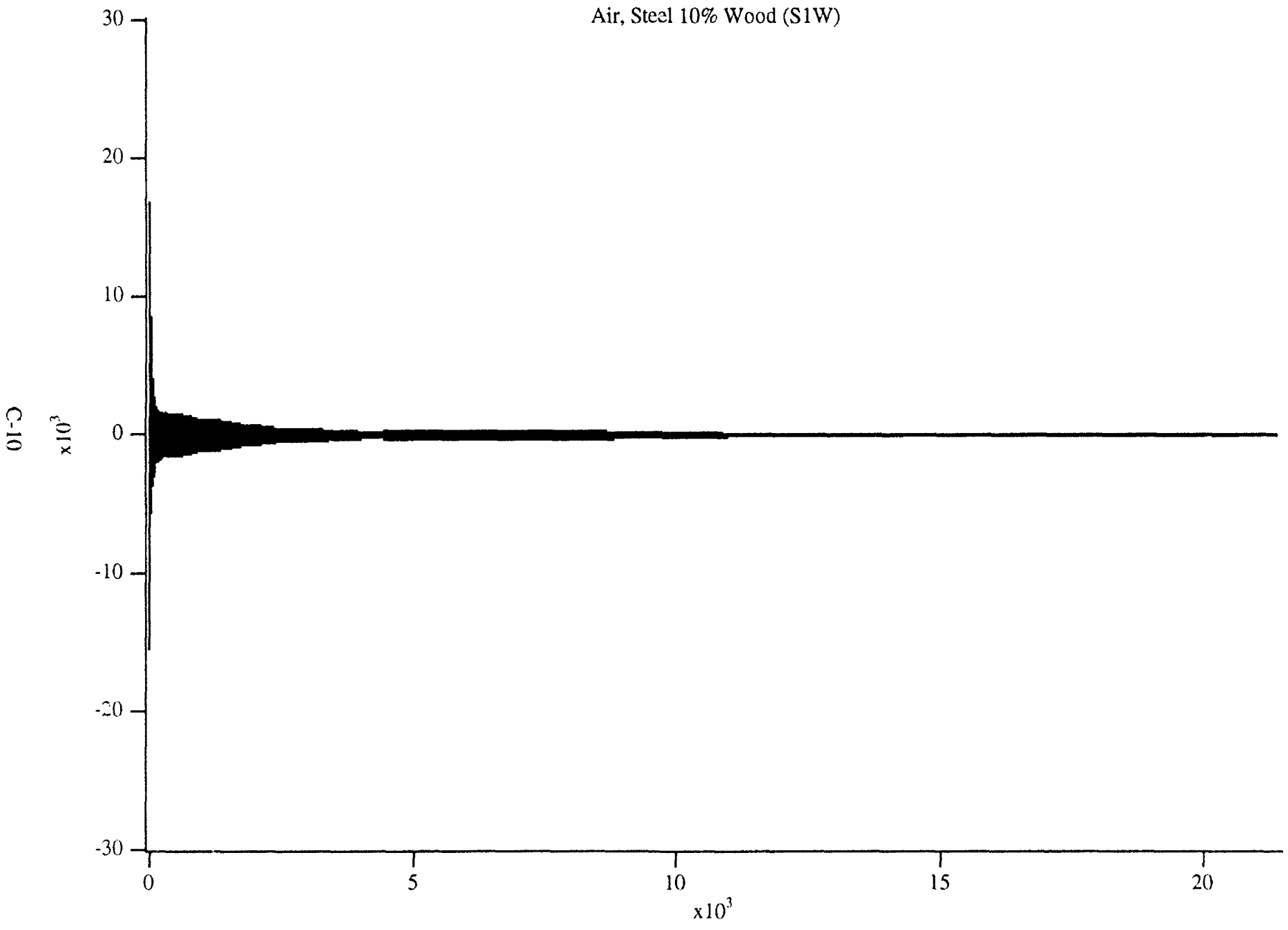
C-5

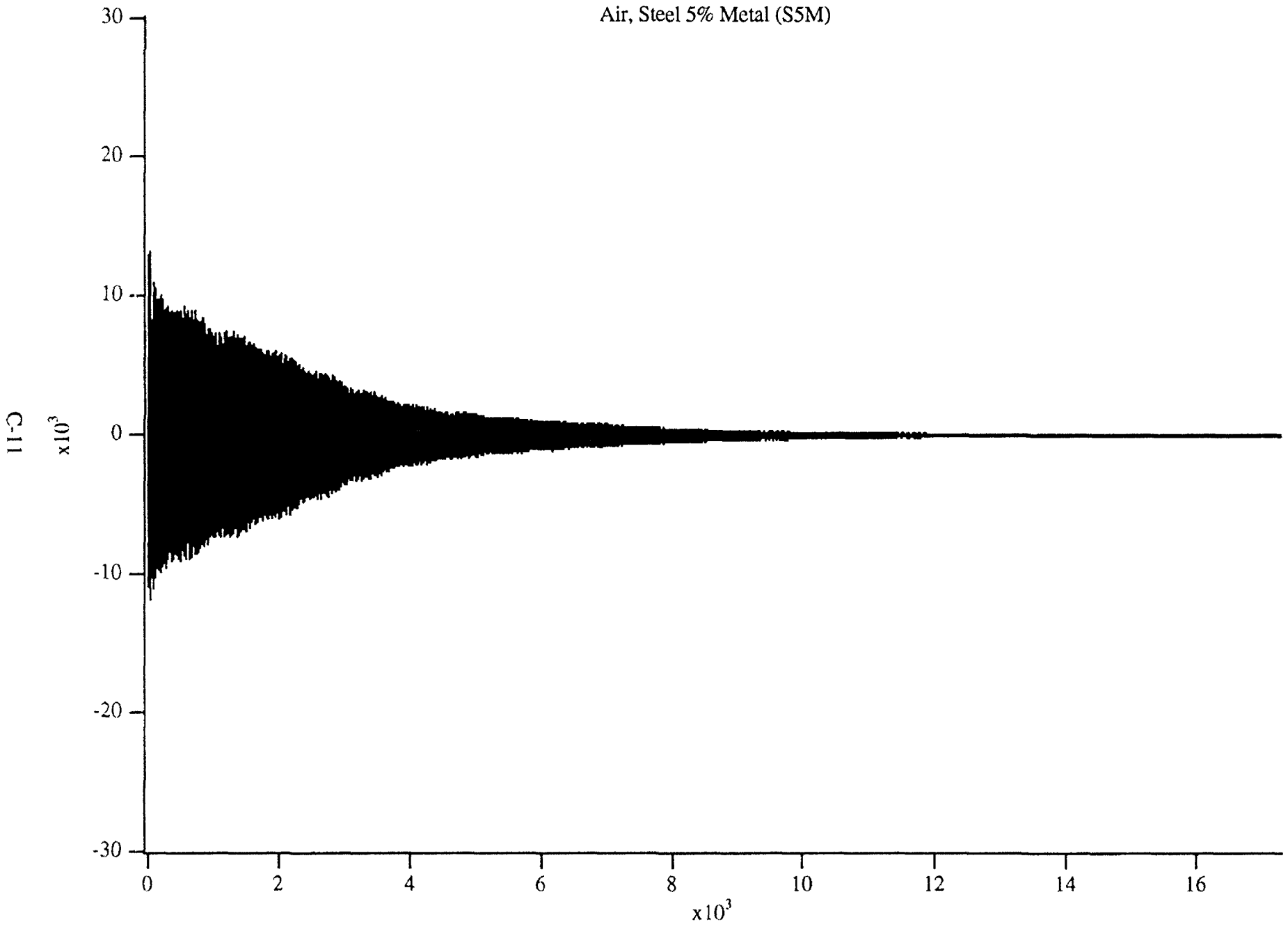

Air, Brass 5% Plastic (B5P)
30
20
10
0
-10
-20
-30
x10^3
0
5
10
15
20
25
30
x10^3

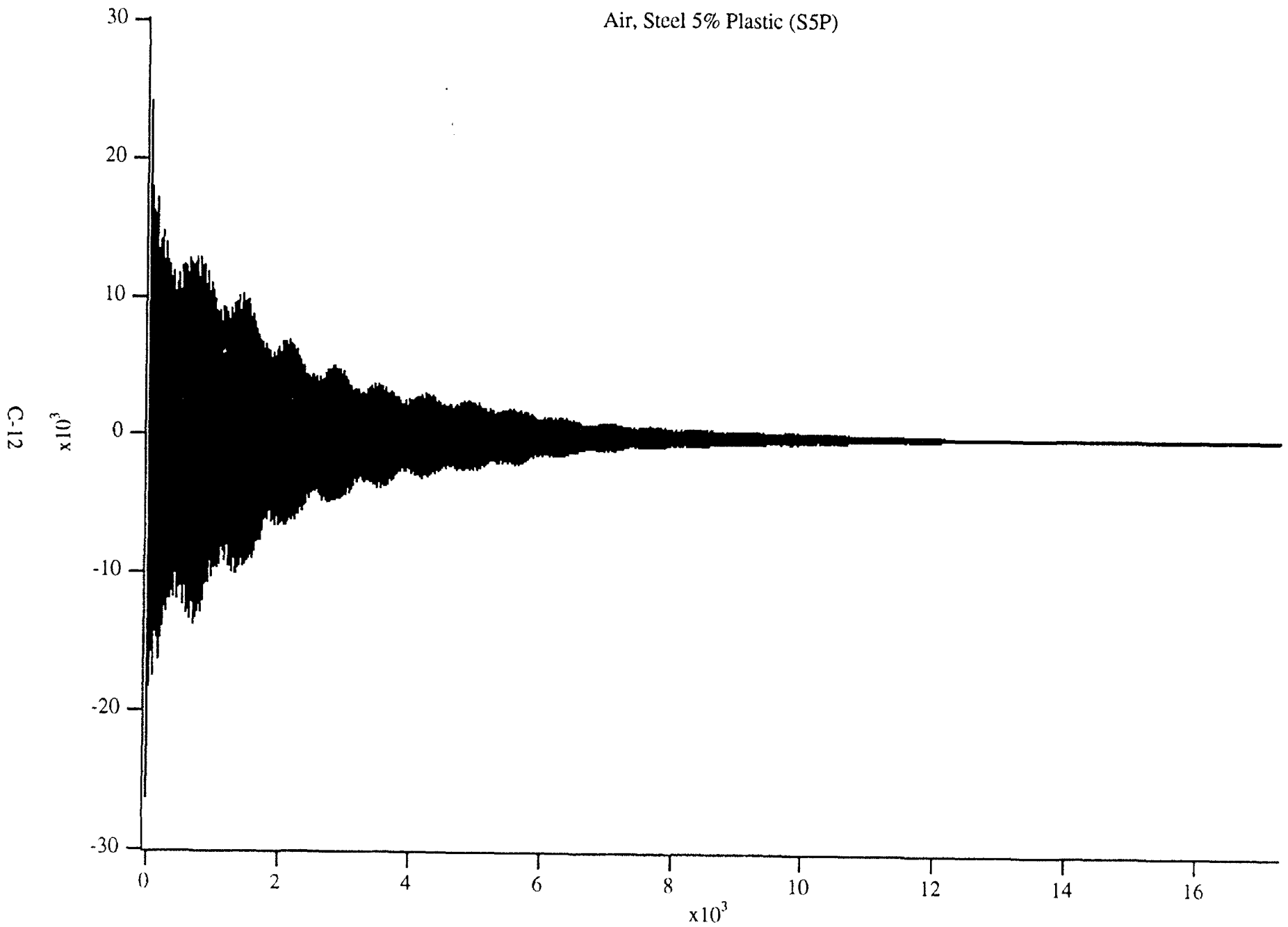

Air, Brass 5% Wood (B5W)
30
20
10
0
-10
-20
-30
x10^3
0
5
10
15
20
25
x10^3

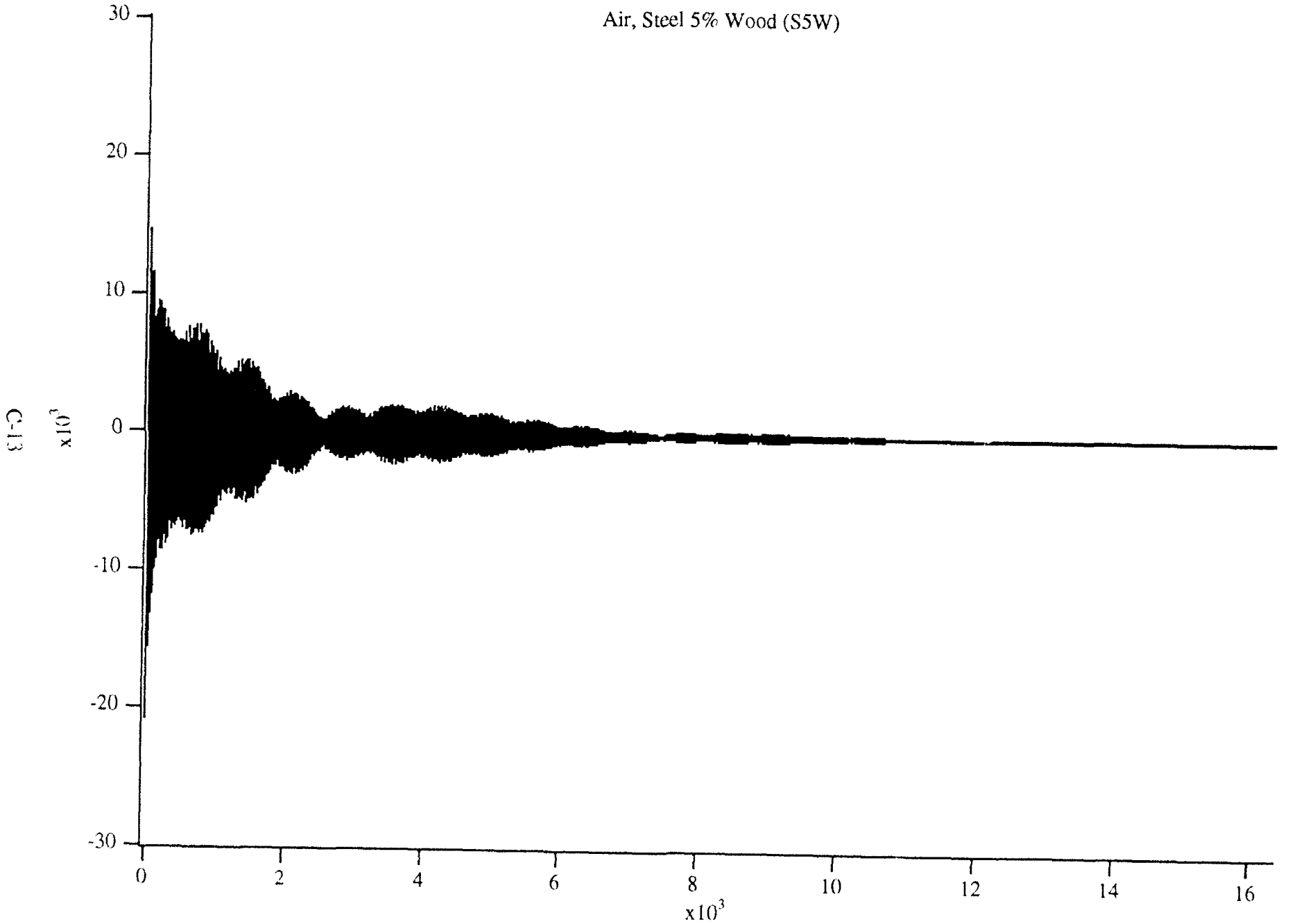

Air, Steel 10% Metal (S1M)
30
20
10
0
-10
-20
-30
x10^3
0
5
10
15
20
25
30
x10^3


Air, Steel 10% Plastic (S1P)
30
20
10
0
-10
-20
-30
x10^3
0
5
10
15
20
x10^3


Air, Steel 10% Wood (S1W)
C-10
x10^3
30
20
10
0
-10
-20
-30
0
5
10
15
20
x10^3


Air, Steel 5% Metal (S5M)
30
20
10
0
-10
-20
-30
x10^3
0
2
4
6
8
10
12
14
16
x10^3


Air, Steel 5% Plastic (S5P)
30
20
10
0
-10
-20
-30
x10^3
0
2
4
6
8
10
12
14
16
x10^3
C-12


Air, Steel 5% Wood (S5W)
30
20
10
0
-10
-20
-30
x10^3
0
2
4
6
8
10
12
14
16
x10^3